THE APPLICATION OF A THEORY

OF MULTI-LEVEL SYSTEMS

TO OPTIMIZATION PROBLEMS

JERRY L. SANDERS

FEBRUARY, 1963

SRC 24-A-63-12

THE APPLICATION OF A THEORY

OF MULTI-LEVEL SYSTEMS

TO OPTIMIZATION PROBLEMS

Jerry L. Sanders

February, 1963

SRC 24-A-63-12

FOREWORD

The Systems Research Center is a research and graduate
study center operating in direct cooperation with all departments
and divisions of Case Institute of Technology. The center brings
together faculty and students in a coordinated program of research
and education in the important techniques of systems theory, develop-
ment, and application.

Research leading to this report was carried on by Mr. Jerry
L. Sanders, Graduate Assistant, under the direction of Dr. Mihajlo
D. Mesarovic, Associate Professor of Engineering at Case and
Director of the Adaptive and Self-Organizing Systems group of the
Systems Research Center.

Raymond J. Nelson, Acting Director
Systems Research Center

ii

ABSTRACT

An attempt is made to simultaneously develop certain
aspects of a theory of multi-level systems and decomposition
techniques for multi-variable optimization problems.

Two level-three goal systems are explored intensively
with particular reference to the method of "coordination" employed
by the second level unit. "Direct" and "Indirect Intervention"
modes are explored. The problem of the quantity of information
required by the second level for the coordination of the first
level systems is investigated. "Direct" and "Indirect Models"
are investigated. The concept of the "Indirect Model" leads to
the development of a Decomposition Principle for Non-Linear
Programming. The concluding chapter presents the analysis of a
numerical example coded for the computer which demonstrates the
utility of certain of the techniques in the context of a specific
application.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 The General Approach

In this thesis we will carry out two parallel lines of development. We will derive some specialized mathematical techniques for the decomposition of large optimization problems. We will, in turn, attempt to place a correspondence between these mathematical results and the processes of decentralized control such as one might find in a decentralized organizational structure. We will also use the process in reverse, e.g. we will consider certain qualitative concepts from the literature on human organizations to suggest avenues of approach for the development of the mathematical aspects. Decentralized decision making systems have been defined by A. Whinston.[22] "By a decentralized decision making system we refer to the following: Given m decisions or actions to be made and n decision makers ($1 < n \leqslant m$), each decision maker is assigned a subset of the m decisions. For the over-all system there is given a criterion function and a space of possible choices involving the m decisions. Each decision maker is assigned a space of possible choices and a criterion function involving at least the decision variables he can partially or totally control."

We hope by adopting this dual point of view to simultaneously develop a framework for the synthesis of "organization-like" structures and at the same time use the mathematical interpretation of these "organizational structures" to develop efficient computational algorithms

for large-scale optimization problems. Our results will have application to engineering control problems[7][t] as well as certain optimization problems which arise in the production theory of the firm.[13][t] We mention in passing the similarity of the problems encountered in these two areas to problems which arise in macro-economic theory. Extensive investigation of control problems in economic theory is currently underway at Case.[19] We will not explore this latter area in any detail.

Large-scale optimization problems have been a constant source of trouble to systems engineering and operations research since their inception. Roughly speaking an optimization problem will be considered "large" when the computational requirements which must be satisfied in order to find the optimal value of the manipulated variables exceeds the capacity of current computing machinery or when the quality of the performance of the system decays significantly in the time required to compute a new control solution.

To our knowledge no previous investigation has been undertaken with the same point of view as adopted here although, of course, we draw on the results of many investigators working in similar areas. Perhaps a combination of two papers have come as close, in spirit at least, to our area of interest as any. These are Whinston (ibid.) and

---

[t]See these references for discussion of optimization problems in these areas.

Dantzig and Wolfe.[8] We quote from the abstract of the latter,
"...Besides holding promise for the efficient computation of large-
scale systems, the principle yields a certain rationale for the
'decentralized decision process' in ⏑ ⏑ theory of the firm....."
The first paper mentioned above is essentially a detailed investi-
gation, both descriptive and analytic, of the consequences of this
statement.

1.2 Types of Problems Considered

We are interested here in "large" multivariable optimization
problems of the type which often occur in the control of complex
systems. Subsequently the reader will often encounter the term
"control problem". This term should be considered to be identical
in meaning to the phrase "...optimization problem which arises in the
optimal control of a complex system...." An exact definition of what
is meant by "control problem" will be given in Chapter 2.

We attempt an "organizational" approach to complex control
problems, i.e. we attempt to break up the over-all control problem
into a number of smaller problems, each of which is to be solved by
a real or fictitious "first level" control unit. In addition we
synthesize one or more "second level" control units whose function is
to coordinate two or more first level controllers. By proceeding in
this way we hope to achieve the following economies: if the process
is a real one, i.e. if the imagined organizational structure can be
realized, then we will enjoy the benefits of parallel operation in

that several parts of the over-all problem will be processed simultan-
eously. If the process is imaginary, i.e. is simply a computational
device, then we have traded the task of solving a large problem for
that of solving a number of smaller ones. In either case this pro-
cedure can lead to significant computational savings as we will see.

It will not always be possible in the models that we examine
to achieve the same quality of performance in the "organizational"
approach as we might have obtained had we been able to solve the
original problem. In the cases where we do not achieve the true
optimum we will attempt to offer a means of computing the difference
in performance between our method and the "true optimum" performance.

The consideration of the best method for breaking up the
original control problem into its respective parts, i.e. "optimal
reticulation", is not one of the problems considered here. This
problem is intimately connected with the concept of "interaction".
No generally accepted definition of this term as yet exists. Exten-
sive work on this and related problems is currently underway at the
Case Systems Research Center.[18] Qualitatively, at least, we would
like to break or "reticulate" the system equations at those points
where the interaction or coupling is weakest. In the problems we
consider we will either assume that the reticulation is given a
priori or we will perform the reticulation more or less arbitrarily,
without regard to the strength of interaction.

2. DEFINITIONS AND PRELIMINARIES

Introduction

In this chapter we will state a rigorous definition of what

we mean by "Control Problem", indicating what information is necessary

in order for a control problem to exist. "Multi-level Control" is

defined and the synthesis procedure which will be investigated in

subsequent chapters is discussed. The level of rigor that the reader

can expect and the exploratory nature of the entire investigation

are explained.

2.1 Definition of the Control Problem

Since we are considering the applicability of our results

to three rather dissimilar fields, we will require a somewhat differ-

ent approach to control problems than has customarily been taken in

the literature.

For our purpose, in order for a control problem to exist,

we require three separate statements.

1. Specification of the utility functional. Here we
   assume that there is specified, a priori, a func-
   tional $g(\cdot, \cdot, \cdots)$ defined over the Cartesian
   product space of the variables $m_1 \cdots m_n$, i.e. over
   $M_1 \times M_2 \times M_3 \cdots M_n$.

2. Specification of Restrictions on the Variables m_i.
   We assume that there is specified in advance a vector
   function ϕ of the variables $m_1 \cdots m_n$ which specify
   the restrictions that the "system" places on the
   problem. Written out these conditions are:

$$\phi_1 (m_1 \cdots m_n) \leq 0$$
$$\vdots$$
$$\phi_j (m_1 \cdots m_n) \leq 0$$

Example 1. For certain engineering problems and for many economic models these restrictions take the form $\phi = 0$ where $\phi = \bar{y} - A\bar{m} - \bar{\epsilon}$. $\bar{y}$ is a n x 1 vector of outputs, A is a n x n matrix, $\bar{m}$ is a n x 1 vector of manipulated variables and $\bar{\epsilon}$ is a n x 1 vector of disturbances.

Example 2. In the production theory of the firm we often encounter the following type of restrictions: $A\bar{m} - \bar{b} \leq 0$ where a typical element of this vector matrix inequality might for example express the fact that the total operating time for any one facility is less than or at most equal to 24 hours in any one day.

3. Specification of the Environment. In general we will assume that the variables $m_1 \ldots m_k$ are controllable by the system or the system manager and that the variables $m_{k+1}, \ldots m_n$ are not under the direct control of the system manager although in some cases these variables may be observable. The specification of the environment will consist of the specification of a probability density function $h(m_{k+1}, \ldots m_n)$ which expresses the joint probability of occurrence of specific numerical values of the external variables. The density $h(m_{k+1}, \ldots m_n)$ may take several forms. It may be a true probability density function or a sub-jective probability statement or a statement of the exact values $m_{k+1}, \ldots m_n$ will take on (in the deterministic case) or the set $\{k+1, \ldots n\}$ may be empty. We will often substitute the vector $\bar{\epsilon}$ for the variables $m_{k+1}, \ldots m_n$ in the chapters that follow.

Now given these three statements we can formulate the complete control problem as follows:

maximize or minimize $\quad E\, g(m_1 \ldots m_k, m_{k+1}, \ldots m_n)$ (2.1)

with respect to $m_1, \ldots m_k$

subject to the restriction $\bar{\phi}\,(m_1 \ldots m_n) \leq 0$

where E is the operation of mathematical expectation

over the variables $m_{k+1} \ldots m_n$.

Example 1

$$\text{minimize} \quad E_{\epsilon} \quad (A\bar{m} + \bar{\epsilon})^T B (A\bar{m} + \bar{\epsilon})$$
$$\text{over } \bar{m}$$

where $\bar{\epsilon}$ is distributed with p.d.f $h(\bar{\epsilon})$.

or stated in the customary way:

$$\text{minimize} \quad E_{\epsilon} y^T B y \quad \text{subject to the constraint}$$
$$\text{over } \bar{m}$$

$$\bar{y} - A\bar{m} - \bar{\epsilon} = 0$$

Example 2

$$\min_{\bar{m}_1, \ldots \bar{m}_n} \quad E_{\epsilon} \sum_{i=1}^{N} \bar{y}_i^T B \bar{y}_i$$

subject to $\bar{y}_1 = \bar{c}$ (a constant vector)

and $\bar{y}_{j+1} = A\bar{y}_j = \bar{m}_{j+1} + \bar{\epsilon}_{j+1}$

where $\bar{\epsilon}_{j+1}$ is distributed with p.d.f $h_{j+1}(\bar{\epsilon}_{j+1})$.

All the subscripts in this example refer to a time index and not to a vector component.

Example 3

$$\text{maximize} \quad \sum_{n=1}^{N} c_n m_n$$

$$\text{subject to} \quad \sum_{j=1}^{P} a_{ij} m_j - b_i \leq 0 \quad i = 1, \ldots \ell$$

where c_n, a_{ij}, and b_i are all fixed constants.

The reader will of course recognize this as the classic linear programming problem.

2.2 Optimal vs. Sub-Optimal Control

The solution to the problem posed in 2.1 will be referred to as the Optimal Control Solution. By solution here we mean the

optimizing values of $m_1 \ldots m_k$, i.e. $m_1^*, \ldots m_k^*$. In the chapters that follow we will occasionally consider ways of generating the controllable variables which do not yield the same values as those generated in the solution to problem 2.1. The solutions will be referred to as Sub-Optimal Control Solutions.

The work "optimal" here is unfortunately a "loaded" word. The solution to 2.1 is the "best that one can do" in the context stated; however, it neglects one critical economic factor—the cost of computation. One might well ask, "What is 'optimal" about a solution that will take 10^7 years to calculate on present computing machinery?" or, "What is 'optimal' about a solution for a chemical processing plant whose expected net profit for the year is $10,000 when the computing machinery required for the computation of the control solution costs on the order of 10^6?" Contrary to being unusual, examples such as those just mentioned are quite prevalent in practice.

Very little work has been done on over-all methods of optimization, i.e. general economic optimization, in control processes. However, the cost of computation considerations indicates that we need to examine certain of our prejudices concerning general distaste for so-called "sub-optimal" solutions, because it is easily possible that these "sub-optimal" procedures are actually economically preferred to the "optimal" solution.

2.3 Single and Multi-Level Control

For the purposes of illustration we will consider some block diagrams of some of the control processes that will appear later. Nearly all of the control processes which occur in engineering and many of those which occur in operations research can be represented as follows:

Fig. 2.1
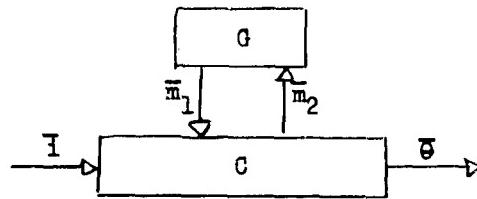
C is the "causal unit" or "system" or "plant". G is the "goal seeking unit" or "controller". This unit is responsible for maintaining a model of the process inherent in C and is assumed to have a utility function. G also has a model of the environment which yields the inputs. The utility function is a function of the inputs $\bar{i}$ the outputs $\bar{\theta}$ and the controllable variables $\bar{m}_1$. The variables $\bar{m}_2$ indicate possible measurements performed by G on C. G is assumed to choose those manipulated variables $\bar{m}_1$ which maximize its utility function subject to the models it has of the environment and the process. We will refer to systems such as Fig. 2.1 as $1\ell - 1g$ systems. This notation refers to the fact that the structure is "single level" and "single goal". By "single level" we mean roughly that no managing or coordinating controllers are present. By "single goal" we mean that only one utility function is involved in the entire system.

Almost the entirety of this thesis will be devoted to the study of the analysis and synthesis of $2\ell - 3\varphi$ systems. Their block diagrams will look either like that in Fig. 2.2 or that in Fig. 2.3.

Fig. 2.2.
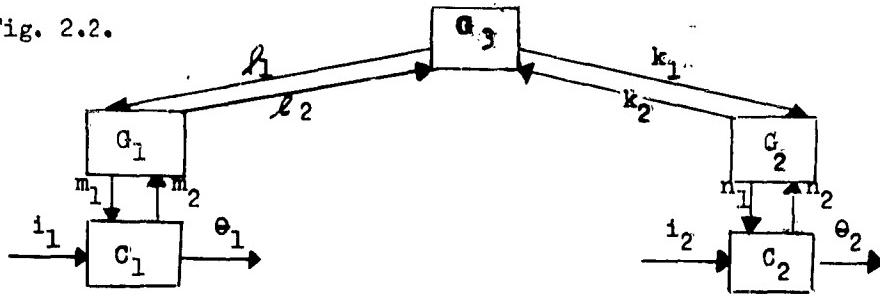
Fig. 2.3.
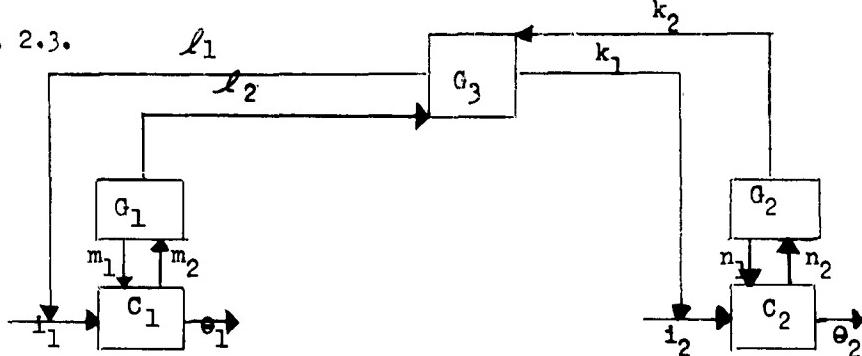
G_3 is, of course, the "coordinating unit" or "managerial unit" or "second level controller". G_3 is entrusted with the utility function of the entire system and with various models of the "behavior" of G_1, C_1, G_2, C_2 and the environment.

The arrows l_2 and k_2 represent the flow of information up from the "first level" systems and l_1 and k_1 represent the flow of control down from G_3. The system represented by Fig. 2.2 is distinguished from 2.3 by the fact that G_3 is employing a method of "coordination" which we will call "indirect intervention" because

we note that G_3's attempts to steer the whole system are made via control-action aimed at the goal-seeking units G_1 and G_2. In Fig. 2.3 we call this "coordination mode" "direct intervention" because the second level's control action is directed toward the causal systems C_1 and C_2.

Example 1 "Indirect Intervention"

Consider a decentralized firm with two operating divisions and a central planning unit where the central planning unit sets the internal or accounting prices of certain commodities such as staff services, etc. In this example the central planning unit plays the role of G_3 and the two operating divisions are represented by (G_1, C_1) and (G_2, C_2). The control actions l_1, and k_1 are the setting of the price of staff services for the respective divisions. The method of coordination here is "indirect" because the central planning unit, by setting the prices, is essentially manipulating parameters in the utility functions of G_1 and G_2.

Example 2 "Direct Intervention"

Consider an international planning agency which is responsible for the economic growth (suitably defined) of two underdeveloped countries. The separate governments are represented by G_1 and G_2. Assume that the planning agency decides to give each country a large sum of money to be immediately injected into the respective economics (C_1, C_2) in the form of construction spending, etc. This method of "control" would be "direct intervention" because direct inputs are injected into the causal systems—in this case, the respective economies.

Obviously, these distinctions are somewhat arbitrary and may or may not be identifiable in any real system. However, we will find these distinctions useful in the chapters to follow. Roughly, we will characterize a goal-seeking unit G as being "second level" if its primary function is the "coordination" of two or more first level units.

By "coordination" we mean the process of the transmission of information and control signals from the second level unit to the first level units for the purpose of improving the over-all performance of the system as measured by the performance functional of the second level unit.

One further component of the over-all problem from the point of view of G_3 is the type of model that this unit has of the structure "below" it. We will classify the model types as either "direct" or "indirect". G_3 is said to have a "direct" model of the system below it if it requires complete knowledge of the structure of (G_1, C_1) and (G_2, C_2) as well as the system's environment. G_3 is said to have an "indirect" model of the system below it if it requires only the knowledge of the system's environment and the utility functions of G_1 and G_2 or certain parameters of these utility functions.

2.4 Remarks concerning Rigor and Intent

The models which appear in the following chapters are in no way to be construed to be final answers concerning the questions posed in this thesis. The intent here is purely exploratory. The amount of room left for investigation is nearly limitless in this area.

In keeping with the exploratory nature of this investigation we will often sacrifice mathematical rigor to the gods of expediency. The analytically inclined reader may be somewhat appalled by the extreme assumptions made concerning the existence of inverses, derivatives, etc.

Summary

In this section we will summarize in outline form the synthesis procedure that we will carry out several times in the succeeding chapters. It should be emphasized that the steps outlined below do not occur in any particular sequence but rather must more or less be carried out simultaneously.

Given the problem:

extremize $\qquad$ $E\ g\ (m_1 \ \dots \ m_n)$
w.r.t. $m_1, \ \dots \ m_k$

subject to the constraints
$\phi(m_1 \ \dots \ m_n) \leq 0$

and an environment prob. density function $\eta(m_{k+1} \ \dots \ m_n)$

1. Decide on a division of control labor between the two first level systems. (Reticulate)

2. Assign objective functionals (utility functions) and constraints to the first level systems.

3. Assign an environment prob. density function to each first level unit.

4. Decide on the model type to be given second level unit.
   a) **Direct Model**
   b) Indirect Model

5. Decide on an intervention method for the second
    level unit.
        a) Direct Intervention
        b) Indirect Intervention

6. Investigate the quality of performance of the
    resultant system.

3. DIRECT INTERVENTION -- STATIC CASE

Introduction

     In this chapter a class of linear, static control problems
are investigated. The synthesis of a $2\ell - 3g$ controller is investi-
gated from the point of view of "Direct Intervention". The second
level unit is assumed to have a "Direct" model of the subsystems
below it.

3.1 Model 1   Deterministic Case

     The purpose of this chapter is to explore several possible
models of control situations where the systems are said to be "static",
i.e. either the control action is applied only once or if the control
action is applied many times it is assumed that the effects of the
applications are statistically independent of each other. All of
the systems in this chapter and the rest of the thesis as well are
assumed to be multi-variable, i.e. several state variables and several
manipulated variables. The system equations are assumed to be
"linear" and constant over time. In mathematical terms this means
that the system acts as a linear operator which maps elements (vectors)
of the "input space" into elements of the "output space". Our
purpose in investigating the use of $2\ell - 3g$ controllers for linear
systems control is not to solve any new problems in the area of the
analysis of linear systems with a small number of state variables,
since these problems have been attacked for many years by many
authors with great success. Our purpose is to "cut our teeth" on

a class of problems that are analytically tractable in the hope
that we will discover methods of approach which will suggest effi-
cient ways of proceeding in very complex control problems, e.g.
large non-linear problems or even linear problems with many state
variables (e.g. more than 20).

Our general approach in these problems will be as follows:
we will take the original multi-variable control problem and attempt
to break it into two or more smaller parts. From these smaller
parts we will construct, generally, two sub-optimal control problems,
each of which neglects the possible interactions with the other
problem. At this point the second level of control will be intro-
duced. It is the purpose of the second level control unit to intro-
duce into the system at an appropriate point a signal which
compensates for the neglected interaction. This chapter attempts
to investigate this method of control in the static case and to
compare its performance with possible alternatives. The first of
these alternatives is the case where the interaction between the
two original "First Level" controllers is completely neglected. The
second alternative, which is, in large complex problems, never
available to us, is that of the true optimal solution.

For simplicity of exposition we will confine ourselves to
operators defined on finite dimensional vector spaces. A theorem
from functional analysis tells us that we can treat the finite
dimensional operators as finitie matrices. (See Appendix A).

<u>Model 1</u> $\overline{y} = A\overline{m} + \overline{\epsilon}$, $|A| \neq 0$ (3.1)

y is an n-component vector $[y_1, \ldots y_n]^T$. T is used to denote the

transpose of a vector or matrix. A is assumed to be an n x n matrix

composed as follows $\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = A$ where A_{11} is a k x k submatrix

(k does not have the same meaning here as in eq. 2.1). A_{12} is

k x (n – k), A_{21} is (n – k) x k and A_{22} is (n – k) x (n – k).

$\overline{m}$ and $\overline{\epsilon}$ are of the same dimensionality as $\overline{y}$. We can view the $\overline{y}$

variables as outputs of the process, $\overline{m}$ as the inputs and the $\overline{\epsilon}$ as

disturbances arising from the environment. In this model the $\overline{\epsilon}$ are

assumed to be deterministic and known. The vectors $\overline{y}$, $\overline{m}$ and $\overline{\epsilon}$ may

either be considered as being composed of numbers or as being composed

of time functions, e.g. $\overline{y} = \begin{bmatrix} y_1(t) \\ \vdots \\ y_n(t) \end{bmatrix}$. The important difference

between the models of this chapter and those of chapter 5 is that

in this chapter $y_i(t + a)$ is independent of $y_j(t)$ for all i, j, t and

a. We assume that a criterion of over-all system performance is given.

It is also assumed that the second level unit is entrusted with the

maximization of this utility. In this model the goal of the system

operator is assumed to be to minimize $\left\{ \overline{y}^T B \overline{y} \right\}$. $\overline{y}^T B \overline{y}$ can be view-
w.r.t. $\overline{m}$

ed as the cost of deviation from a fixed operating point which in

this case is 0. The operation of the system around any predescribed

operating point can be accommodated by simply changing the origin of

coordinates in the space in which $\overline{y}$ is measured. In other words,

the job of the system operator in this case is to keep the system as

close to the fixed operating point as possible. Certain disturbances $\vec{\epsilon}$ impinge on the system from the environment but the operator can change the vector $\bar{m}$ in an attempt to offset these disturbances. The cost of deviating from the preset optimum conditions is roughly proportional to the square of the deviation since $\bar{y}^T B \bar{y}$ is a quadratic form. B is composed as follows $B = \begin{bmatrix} B_{11} & 0 \\ 0 & B_{22} \end{bmatrix}$ where B_{11} is $k \times k$ and B_{22} is $(n - k) \times (n - k)$. (B_{ii} is assumed to be positive definite). At this point we introduce some notation which we will use throughout. We let $g = \bar{y}^T B \bar{y}$ and $g* = \min\limits_{\bar{m}} \left\{ \bar{y}^T B \bar{y} \right\}$. g then is the actual performance and $g*$ is the optimal performance. We will now assume that the problem of determining the optimal $\bar{m}$ is too large a problem to be solved by current optimization techniques. We will therefore subdivide the problem as previously described. We let the subvector $\begin{bmatrix} y_1 & \cdots & y_k \end{bmatrix}^T = \bar{y}^1$. Similarly $\begin{bmatrix} y_{k+1} & \cdots & y_n \end{bmatrix}^T = \bar{y}^2$, $\begin{bmatrix} m_1 & \cdots & m_k \end{bmatrix}^T = \bar{m}^1$, $\begin{bmatrix} m_{k+1} & \cdots & m_n \end{bmatrix}^T = \bar{m}^2$, $\begin{bmatrix} \epsilon_1 & \cdots & \epsilon_k \end{bmatrix}^T = \bar{\epsilon}^1$, and $\begin{bmatrix} \epsilon_{k+1} & \cdots & \epsilon_n \end{bmatrix}^T = \bar{\epsilon}^2$.

We now turn the responsibility for $\bar{y}^1$ and $\bar{m}^1$ over to a controller called "First Level System#1" and $\bar{y}^2$ and $\bar{m}^2$ over to "First Level System #2". System 1 is set the following task:

$$\min_{\left\{\bar{m}^1\right\}} \quad {}_1\bar{y}^{1T} B_{11} \; {}_1\bar{y}^1 \qquad \text{subject to} \qquad (3.2)$$

$${}_1\bar{y}^1 = A_{11} \, \bar{m}^1 + \bar{\epsilon}^1$$

The extra subscript introduced here is to denote, for example, that ${}_1\bar{y}^1$ is System 1's "model" of the vector $\bar{y}^1$. The reader will note

that this model is actually incorrect from the point of view of the original problem posed. System 1 is neglecting the element A_{12} which is a cross-coupling term coupling $\bar{y}^1$ to $\bar{m}^2$. System 1 neglects this term on purpose in order to separate its problem from System 2's.

System 2's problem is to

$$\text{minimize} \qquad _2\bar{y}^{2^T} \ B_{22} \ _2\bar{y}^2 \qquad\qquad (3.3)$$

$$\text{subject to} \qquad _2\bar{y}^2 = A_{22} \ \bar{m}^2 + \bar{\epsilon}^2$$

The solution to problems (3.2) and (3.3) can easily be derived (see Appendix A). Equating the vector derivative of the quadratic forms in (3.2) and (3.3) to zero we have

$$A_{22}^T \ B_{22} \ (A_{22} \ \bar{m}^2 + \bar{\epsilon}^2) = 0 \qquad\qquad (3.4)$$

$$\text{and} \qquad A_{11}^T \ B_{11} \ (A_{11} \ \bar{m}^1 + \bar{\epsilon}^1) = 0 \qquad\qquad (3.5)$$

We assume (as always, for the sake of simplicity) that B_{11} and A_{11} are non-singular and that $A_{11}^T \ B_{11} \ A_{11}$ is positive definite. Therefore, we have

$$\bar{m}^{1*} = -A_{11}^{-1} \ \bar{\epsilon}^1 \qquad \text{and} \qquad \bar{m}^{2*} = -A_{22}^{-1} \ \bar{\epsilon}^2.$$

We note that the $\bar{\epsilon}$'s are assumed to be known to all the systems involved in the control of the process. If these vectors $\bar{m}^{i*}$ were allowed to be implemented as they are the actual outputs of the process would be

$$\bar{y}^1 = A_{11} \ \bar{m}^{1*} + A_{12} \ \bar{m}^{2*} + \bar{\epsilon}^1 = -A_{12} \ A_{22}^{-1} \ \bar{\epsilon}^2 \qquad\qquad (3.6)$$

$$\bar{y}^2 = A_{21} \ \bar{m}^{1*} + A_{22} \ m^{2*} + \bar{\epsilon}^2 = -A_{21} \ A_{11}^{-1} \ \bar{\epsilon}^1$$

At this point we will introduce the second level controller. System #3. System 3 has the responsibility for the control of the entire process and as such it will be concerned with the complete model of the system and with the system performance functional. Specifically its task will be to insert appropriate inputs into the system to compensate for the interaction neglected by the first level units. To be more precise, we will assume that System 3 has two optimization problems.

Problem 1   minimize $_3\overline{y}^{1^T} B_{11}\ _3\overline{y}^{1}$ (3.7)

subject to $_3\overline{y}^{1} = A_{11}\ \overline{m}^{1} + A_{12}\ \overline{m}^{2} + A_{12}\ \overline{\eta}^{2} + \overline{\epsilon}^{1}$

where $\overline{\eta}^{2}$ is a k-component vector manipulated by the second level unit.

Problem 2   minimize $_3\overline{y}^{2^T} B_{22}\ _3\overline{y}^{2}$ (3.8)

subject to $_3\overline{y}^{2} = A_{21}\ \overline{m}^{1} + A_{22}\ \overline{m}^{2} + A_{21}\ \overline{\eta}^{1} + \overline{\epsilon}^{2}$

$\overline{\eta}^{1}$ is an (n - k) component vector manipulated by the second level unit.

We have already assumed that System 3 has the knowledge of the complete structure of the system itself as well as the optimal control vectors from the first level systems, i.e. $\overline{m}^{1*}$ and $\overline{m}^{2*}$. The first level units are required only to have knowledge of their own subsystems and their part of the system goal in addition to the relevant disturbance sub-vector.

Now $_3\bar{y}^1 = -A_{12} A_{22}^{-1} \bar{\epsilon}^2 + A_{12} \bar{\eta}^2$ and

$$_3\bar{y}^2 = -A_{21} A_{11}^{-1} \bar{\epsilon}^1 + A_{21} \bar{\eta}^1.$$

The solution to System 3's two problems is now obvious. It should set $\bar{\eta}^{2*} = A_{22}^{-1} \bar{\epsilon}^2$ and $\bar{\eta}^{1*} = A_{11}^{-1} \bar{\epsilon}^1$. Therefore $_3\bar{y}^1 = _3\bar{y}^2 = 0$ and the value of the performance functional for the entire system, i.e. $\bar{y}^T B \bar{y}$ takes on the value 0.

Let us examine briefly the implications of the results of this rather elementary example. First notice that the system performance functional $\bar{y}^T B \bar{y}$ assumes its absolute minimum. This means that no other control algorithm will produce better result; for this system. This is in fact then optimal control. We have taken the original problem apart and recombined it along organizational lines in such a way as to get the best possible performance. A reasonable question at this point is "Why?" The reason here and for all subsequent models is computational. In nearly all real multi-variable control problems the problem of computing a sequence of optimal control variables meets with many computational difficulties. Assume that the matrix A is of even dimension. If we had tried to solve the original optimization problem directly we would have re-quired the inverse of the n dimension A matrix. In our problem we required only the inverse $(k = \frac{n}{2})$ of matrices of size $\frac{n}{2}$. The time for the computation of the inverse of a matrix on a digital computer goes up roughly as the cube of the matrix size. Here and particularly in the dynamic case, as we will see later, this kind of

difference can mean the difference between being able to solve the problem and being forced to neglect certain potentially important aspects of the performance.

The reader can undoubtedly level several objections at this model. The first might be that this might not be a feasible approach, i.e. to intervene in the performance of the system in this way, and this we would readily admit. Secondly, one might argue that we are not considering the cost of the manipulated variables. This will be considered in subsequent models. Thirdly, one might say that linear models of this kind are so naive in structure and formalism that they can never represent the control of a real process. This allegation has a certain amount of truth to it and may recur in several of the subsequent models. We will attempt to answer it here and offer no further apologies later. First of all, it is not strictly correct, since models of similar type have been employed in both engineering and economics for many years with varying degrees of success. The majority of models encountered in operations research, however, are extremely non-linear. Unfortunately, analytical results are almost completely lacking for the control solutions to non-linear systems. Therefore, whether we like linear models or not we are forced to them in order to try our approach on a fairly broad class of system models. The digital computer offers us very little succor here, because it yields only one solution at a time and not a class of solutions as we require.

The interesting feature of this model was that we were able to achieve the true optimum with our method of control system synthesis. As we will see subsequently, this is not always possible with this method.

3.2 Model 2   Probabilistic Case

This model is to be formulated in exactly the same way as Model 1, except that the disturbance vector $\overline{\epsilon}$ is assumed to be a vector of random variables. The first moments of all the components are assumed known and finite, i.e. $E\overline{\epsilon} = \overline{\mu} = \begin{bmatrix} E\epsilon_1 \\ \vdots \\ E\epsilon_n \end{bmatrix}$. All the second

order moments of $\overline{\epsilon}$ are assumed finite. Now we can no longer use functionals of the type $\overline{y}^T B \overline{y}$ as an optimization criterion because these functionals themselves become random quantities; however, we can use their expectation as an extremization criterion.

As before, then,

$$\begin{bmatrix} \overline{y}^1 \\ \overline{y}^2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \overline{m}^1 \\ \overline{m}^2 \end{bmatrix} + \begin{bmatrix} \overline{\epsilon}^1 \\ \overline{\epsilon}^2 \end{bmatrix} \tag{3.9}$$

is our original system model.

System 1's control problem, namely,

$$g_1^* = \min_{\overline{m}^1} \; E\left\{ {}_1\overline{y}^{1^T} B_{11} \; {}_1\overline{y}^1 \right\} \qquad \text{subject to} \tag{3.10}$$

$${}_1\overline{y}^1 = A_{11} \; \overline{m}^1 + \overline{\epsilon}^1 \;,$$

has this solution $\quad \overline{m}^{1^*} = -A_{11}^{-1} \; \overline{\mu}^1, \quad \overline{\mu}^1 = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_k \end{bmatrix}.$

For System 2 we have

$$g_2^* = \min_{\overline{m}^2} \quad E\left\{ {}_2\overline{y}^{2^T} B_{22} \, {}_2\overline{y}^2 \right\} \quad \text{subject to} \quad (3.11)$$

$$_2\overline{y}^2 = A_{22} \, \overline{m}^2 + \overline{\epsilon}^2 \quad \text{and} \quad \overline{m}^{2*} = -A_{22}\overline{\mu}^2.$$

Therefore, $\quad \overline{y}^1 = A_{11}\left(-A_{11}^{-1}\overline{\mu}^1\right) + A_{12}\left(-A_{22}^{-1}\overline{\mu}^2\right) + \overline{\epsilon}^1 \qquad (3.12)$

$$\overline{y}^2 = A_{21}\left(-A_{11}^{-1}\overline{\mu}^1\right) + A_{22}\left(-A_{22}^{-1}\overline{\mu}^2\right) + \overline{\epsilon}^2$$

$$\overline{y}^1 = \overline{\epsilon}^1 - \overline{\mu}^1 - A_{12} A_{22}^{-1}\overline{\mu}^2$$

$$\overline{y}^2 = \overline{\epsilon}^2 - \overline{\mu}^2 - A_{21} A_{11}^{-1}\overline{\mu}^1$$

Now for System 3 we have

$$g_3^* = \min_{\overline{\eta}^1,\overline{\eta}^2} \quad E\left\{ {}_3\overline{y}^{1^T} B_{11} \, {}_3\overline{y}^1 + {}_3\overline{y}^{2^T} B_{22} \, {}_3\overline{y}^2 \right\} \qquad (3.13)$$

subject to $\quad {}_3\overline{y}^1 = A_{11}\overline{m}^{1*} + A_{12}\overline{m}^{2*} + A_{12}\,\overline{\eta}^2 + \overline{\epsilon}^1 \qquad (3.14)$

$$_3\overline{y}^2 = A_{21}\overline{m}^{1*} + A_{22}\overline{m}^{2*} + A_{21}\,\overline{\eta}^1 + \overline{\epsilon}^2$$

The solution to (3.13) yields

$$\overline{\eta}^{1*} = A_{11}^{-1}\overline{\mu}^1 \qquad (3.15)$$

$$\overline{\eta}^{2*} = A_{22}^{-1}\overline{\mu}^2$$

$$_3\overline{y}^1 = \overline{\epsilon}^1 - \overline{\mu}^1 \qquad (3.16)$$

$$_3\overline{y}^2 = \overline{\epsilon}^2 - \overline{\mu}^2$$

Let us check this result against what we might have obtained by solving the "true optimum" problem:

$$g* = \min_{\overline{m}} \quad E\left(\overline{y}^T B \overline{y}\right) \qquad (3.17)$$

subject to $\overline{y} = A\overline{m} + \overline{\epsilon}$

The expectation and derivative of g yields

$$A^T B \left(A\overline{m} + \overline{\mu}\right) = 0 \qquad \text{if } A^T B A \text{ is positive definite and}$$

B and A have inverses then $\overline{m}* = -A^{-1} \overline{\mu}$

hence $\quad \overline{y} = \overline{\epsilon} - \overline{\mu} = \begin{bmatrix} \overline{\epsilon}^1 & - & \overline{\mu}^1 \\ \overline{\epsilon}^2 & - & \overline{\mu}^2 \end{bmatrix}$ and we can see that this result

is identical to (3.16).

3.3 Model 3   Cost of Control

In this model we consider a situation similar to those in Models 1 and 2, except that we now consider an extra term in the performance functionals which represents the cost of the manipulated variables $\overline{m}$.

System 1

$$g_1^* = \min_{\overline{m}^1} \quad \left\{ {}_1\overline{y}^{1T} B_{11} \, {}_1\overline{y}^1 + \overline{m}^{1T} C_{11} \overline{m}^1 \right\} \qquad (3.18)$$

${}_1\overline{y}^1 = A_{11} \overline{m}^1 + \overline{\epsilon}^1, \overline{\epsilon}^1$ are assumed to be deterministic and known to all subsystems in this model. Here as in all subsequent models the necessary conditions for the existence of minima, maxima and inverses will, a priori, be assumed to exist. For the model above $\overline{m}^{1*} = - \left[A_{11}^T B_{11} A_{11} + C_{11} \right]^{-1} A_{11}^T B_{11} \overline{\epsilon}^1$.

Similarly for System 2:

$$g_2^* = \min_{\overline{m}^2} \left\{ {}_2\overline{y}^{2^T} B_{22}\, {}_2\overline{y}^2 + \overline{m}^{2^T} C_{22}\, \overline{m}^2 \right\} \qquad (3.19)$$

$$\overline{m}^{2*} = -\left[A_{22}^T B_{22} A_{22} + C_{22} \right]^{-1} A_{22} B_{22}\, \overline{\epsilon}^2$$

System 3

$$g_3^* = \min_{\overline{\eta}^1,\, \overline{\eta}^2} \left\{ {}_3\overline{y}^{1^T} B_{11}\, {}_3\overline{y}^1 + (\overline{m}^1 + \overline{\eta}^1)^T C_{11} (\overline{m}^1 + \overline{\eta}^1) \right. \qquad (3.20)$$

$$\left. + {}_3\overline{y}^{2^T} B_{22}\, {}_3\overline{y}^2 + (\overline{m}^2 + \overline{\eta}^2)^T C_{22} (\overline{m}^2 + \overline{\eta}^2) \right\}$$

where ${}_3\overline{y}^1$ and ${}_3\overline{y}^2$ are defined as in (3.14) with the $\overline{m}^{i*}$ given

by (3.19). Differentiation of (3.20) yields

$$\overline{\eta}^{1*} = -\left[A_{21}^T B_{22} A_{21} + C_{11} \right]^{-1} \left\{ A_{21}^T B_{22} \left[A_{21}\, \overline{m}^{1*} + A_{22}\, \overline{m}^{2*} \right. \right. \qquad (3.21)$$

$$\left. \left. + \overline{\epsilon}^2 \right] + C_{11}\, \overline{m}^{1*} \right\}$$

$$\overline{\eta}^{2*} = -\left[A_{12}^T B_{11} A_{12} + C_{22} \right]^{-1} \left\{ A_{12}^T B_{11} \left[A_{12}\, \overline{m}^{2*} + A_{11}\, \overline{m}^{1*} \right. \right.$$

$$\left. \left. + \overline{\epsilon}^1 \right] + C_{11}\, \overline{m}^{2*} \right\}$$

The problem of comparing this mode of control with "true optimal"
control is not without its difficulties. It is logically possible
for this mode to come out better than the "true optimal" control
because of the added terms in ${}_3\overline{y}^i$ of $A_{12}\, \overline{\eta}^2$ and $A_{21}\, \overline{\eta}^1$. The
submatrices A_{12} and A_{21} offer amplification of the control
vectors $\overline{\eta}^i$ in a way which is not available to the "true optimal"
formulation. If we are to charge ourselves for this amplification

then perhaps the appropriate goal functional is

$$\jmath = \bar{y}^T \; B \; \bar{y} + (\bar{\chi} + \bar{m})^T \; C \; (\bar{\chi} + \bar{m}) \qquad \text{where} \; \bar{\chi} = \begin{bmatrix} A_{21} & \bar{\eta}^1 \\ A_{12} & \bar{\eta}^2 \end{bmatrix}.$$

In any event it appears that in order to perform any quantitative comparison of these two different approaches one would be required to perform a numerical experiment on a computer to determine the sign and magnitude of the performance difference.

We now consider the comparison of this system with an alternative which is the same as our Direct Intervention case, except that the second level unit is missing. Hence nothing is compensating for the neglected interaction between the first level systems. We will call this the non-interacting case. This system has exactly the same performance functional except that, of course, the terms involving $\bar{\eta}^i$ will be missing. We note that the $\bar{m}^{-i*}$ are the same for each case. As a result of these observations we can easily state and prove the following theorem.

Theorem: Let $g_A^* = \min \left[\bar{y}_A^T \; B \; \bar{y}_A + m_A^T \; C \; m_A \right]$

$$g_B^* = \min \left[\bar{y}_B^T \; B \; \bar{y}_B + (m_A + \eta)^T \; C \; (m_A + \eta) \right]$$

where $\bar{y}_B = \bar{y}_A + A* \eta$ $\qquad\qquad$ where $A* = \begin{bmatrix} 0 & A_{12} \\ A_{21} & 0 \end{bmatrix}$

then $g_B^* \leq g_A^*$

Proof: First of all η was chosen in such a way that g_B is minimized. $\eta = 0$ is an admissible control vector. If $\eta = 0$ $g_B^* = g_A^*$.

Suppose $\exists$ an $\eta* \neq 0$ which minimized g_B and $g_B (\eta*) > g_A*$ but clearly $g_B (0) < g_B (\eta*)$ contrary to the assumed existence of $\eta*$ ∴ either $(g_B* < g_A*)$ or $(\eta = 0$ and $g_B* \leq g_A*)$.

We conclude from this theorem then that in the case where we are required to pay for the variables that we manipulate that we can always do at least as well as if we neglect the interaction between the two first level systems.

Similar analysis could easily be carried for the case where $\bar{\epsilon}$ is a random vector.

Summary - - Conclusions and Applications

We began in this chapter by introducing a method of multi-level control called Direct Intervention. In reference to the modeling considerations introduced in Chapter 2 it should be remarked that in this chapter the second level unit is employing what we called a "Direct" model of the first level units.

We began by considering an over-all control problem of the form:

1. $g* = \min \bar{y}^T B \bar{y}$ where $\bar{y} = A\bar{m} + \bar{\epsilon}$
$\bar{\epsilon}$ was assumed to be known and B to be block diagonal. Here we saw that the multi-level approach yielded the optimum solution.

2. Next we consider

$g* = \min E \left\{ \overline{y}^T B \overline{y} \right\}$ where $\overline{y} = A\overline{m} + \overline{\epsilon}$
where here $\overline{\epsilon}$ is a stochastic vector with known
mean. Again we found that if B were block diagonal
we could achieve the optimum solution.

3. Here we had

$g* = \min \overline{y}^T B \overline{y} + \overline{m}^T C \overline{m}$
subject to $\overline{y} = A\overline{m} + \overline{\epsilon}$ where B and C are
block diagonal and $\overline{\epsilon}$ is a known disturbance. We
discussed certain difficulties of comparison with the
"true optimal" case, but we were able to show that
this mode of "intervention" always provided a better
control solution than one could achieve by neglect-
ing the cross-coupling terms.

4. INDIRECT INTERVENTION - STATIC CASE

Introduction

In this chapter our aim is to explore a set of control models quite similar to those in the previous chapter. The principal difference will be the method of coordination employed by the second level unit. The previous chapter was concerned with what we might call "direct signal intervention" models. In that case, the second level unit was supposed to be able either directly or indirectly to insert extra inputs into the system at appropriate points. In this chapter the second level unit influences the outputs of the system by setting the values of certain parameters in the performance functionals of the first level systems. As we will see, this method has both its advantages and disadvantages. Among its advantages are the following: It is possible to effect large reductions in the necessary computing time and memory required to compute the control solutions. Secondly, it virtually can be applied to all classes of control problems—at least all those for which it is possible to find a method of reticulation. For example, it is difficult to conjure up a model used in the area of Operations Research where the direct intervention approach could be applied; however, the method we are about to describe can be applied in several situations rather naturally (see, for example, Chapter 8).

Among the disadvantages of this approach is that a complete closed-form solution is somewhat more difficult than the

models of Chapter 3. Secondly, it generally reduces the controllability (see Appendix D) of the system and this can hurt the performance (this problem will be discussed in more detail in section 4.3) of the system as measured by the performance criterion of the second level unit.

In short, models to be discussed in this chapter are inspired by an attempt to trade off controllability and possibly performance quality against the cost of calculating the optimal control vectors.

4.1 Model 1

Again we start with two first level units.

System 1.
$$_1\bar{y}^1 = A_{11}\ \bar{m}^1 + \bar{\epsilon}^1$$

$$g_1^* = \min_{\bar{m}^1}\quad E\left\{_1\bar{y}^{1^T}B_{11}\ _1\bar{y}^1 + 2\beta_1\ \bar{1}\cdot\bar{m}^1\right\} \qquad (4.1)$$

where $_1\bar{y}^1$, $\bar{m}^1$ and $\bar{\epsilon}^1$ are defined as in Chapter 3. The $\bar{\epsilon}$'s are assumed to have fixed, known means and finite second moments. β_1 is a scalar parameter controlled by the second level unit. It is also assumed that the first level units do not anticipate the values of β given by the second level unit, i.e. they wait for the values of β to be transmitted before the values of the manipulated variables are set. $\bar{1}$ is the vector $\begin{bmatrix} 1, & 1, & \dots & 1 \end{bmatrix}$.

Similarly for System 2:

$$2\bar{y}^2 = A_{22}\,\bar{m}^2 + \bar{\epsilon}^2$$

$$g_2{}^* = \min_{\bar{m}^2}\ E\left\{2\bar{y}^{2^T} B_{22}\ 2\bar{y}^2 + 2\beta_2\ \bar{1}\cdot\bar{m}^2\right\} \tag{4.2}$$

The optimal control vectors for the first level units are

$$\bar{m}^{1*} = -\left[A_{11}^T B_{11} A_{11}\right]^{-1}\left\{\beta_1\cdot I + A_{11}^T B_{11}\ \bar{\mu}^{-1}\right\} \tag{4.3}$$

$$\bar{m}^{2*} = -\left[A_{22}^T B_{22} A_{22}\right]^{-1}\left\{\beta_2\cdot\bar{1} + A_{22}^T B_{22}\ \bar{\mu}^{-2}\right\}$$

where $\bar{\mu}^{-1} = E\left(\bar{\epsilon}^1\right)$.

Now for the second level unit we pose the task of manipulating the parameters β_1 and β_2 in such a way that the neglected interaction (via A_{12}, A_{21}) is at least partially compensated.

We let

$$g_3{}^* = \min_{\beta_1,\beta_2}\ E\left\{\begin{pmatrix}\bar{y}^1\\\bar{y}^2\end{pmatrix}^T B\begin{pmatrix}\bar{y}^1\\\bar{y}^2\end{pmatrix} + \bar{C}^T\begin{pmatrix}\bar{m}^1\\\bar{m}^2\end{pmatrix}\right\} \tag{4.4}$$

$\bar{C}$ is a vector of fixed quantities.

$$\bar{y}^1 = A_{11}\,\bar{m}^{1*} + A_{12}\,\bar{m}^{2*} + \bar{\epsilon}^1 \tag{4.5}$$

$$\bar{y}^2 = A_{21}\,\bar{m}^{1*} + A_{22}\,\bar{m}^{2*} + \bar{\epsilon}^2$$

using (4.3) with (4.4) and (4.5) we have

$$g_3^* = \min_{\beta_1,\beta_2} E\left\{ \left[\beta_1 \phi_{11} + \beta_2 \phi_{12} + \psi_{11} + \psi_{12} + \bar{\epsilon}^1 \right]^T \right. \tag{4.6}$$

$$B_{11} \left[\beta_1 \phi_{11} + \beta_2 \phi_{12} + \psi_{11} + \psi_{12} + \bar{\epsilon}^1 \right]$$

$$+ 2 \left[\beta_1 \phi_{11} + \beta_2 \phi_{12} + \psi_{11} + \psi_{12} + \bar{\epsilon}_1 \right] B_{12} \left[\beta_1 \phi_{21} + \beta_2 \phi_{22} \right.$$

$$\left. + \psi_{21} + \psi_{22} + \bar{\epsilon}_2 \right]$$

$$+ \left[\beta_1 \phi_{21} + \beta_2 \phi_{22} + \psi_{21} + \psi_{22} + \bar{\epsilon}_2 \right]^T B_{22} \left[\beta_1 \phi_{21} + \beta_2 \phi_{22} \right.$$

$$\left. + \psi_{21} + \psi_{22} + \bar{\epsilon}_2 \right]$$

$$- \beta_1 \bar{c}_1 \cdot \left[A_{11}^T B_{11} A_{11} \right]^{-1} \cdot \bar{1} - \bar{c}_1 \left[A_{11}^T B_{11} A_{11} \right]^{-1} A_{11}^T B_{11} \mu^{-2}$$

$$\left. - \beta_2 \bar{c}_2 \cdot \left[A_{22}^T B_{22} A_{22} \right]^{-1} \cdot \bar{1} - \bar{c}_2 \left[A_{22}^T B_{22} A_{22} \right]^{-1} A_{22}^T B_{22} \mu^{-2} \right\}$$

$$\text{where } \bar{c}_1 = \left[c_1 \cdots c_k \right], \quad \bar{c}_2 = \left[c_{k+1} \cdots c_n \right]$$

$$\phi_{11} = A_{11} \left[A_{11}^T B_{11} A_{11} \right]^{-1} \cdot \bar{1} \qquad \psi_{11} = A_{11} \left[A_{11}^T B_{11} A_{11} \right]^{-1} A_{11} B_{11} \mu^{-1}$$

$$= B_{11}^{-1} A_{11}^{T^{-1}} \cdot \bar{1} \qquad\qquad = \mu^{-1}$$

$$\phi_{12} = A_{12} \left[A_{22}^T B_{22} A_{22} \right]^{-1} \cdot \bar{1} \qquad \psi_{12} = A_{12} A_{22}^{-1} \mu^{-2}$$

$$\phi_{22} = B_{22}^{-1} A_{22}^{-1^T} \cdot \bar{1} \qquad\qquad \psi_{22} = \mu^{-2}$$

$$\phi_{21} = A_{21} \left[A_{11}^T B_{11} A_{11} \right]^{-1} \cdot \bar{1} \qquad \psi_{21} = A_{21} A_{11}^{-1} \mu^{-1}$$

B is not necessarily block diagonal in this model but it is assumed symmetric and positive definite. After some painful but simple manipulations, one could take the expectation and the derivative of (4.6) with respect to β_1 and β_2 and equate them to zero to

find the optimizing values of β_1 and β_2. A brief inspection of
(4.6) shows that one would get two equations of the following form:

$$a_1 \; \beta_1* + a_2 \; \beta_2* + a_3 = 0 \tag{4.7}$$

$$b_1 \; \beta_1* + b_2 \; \beta_2* + b_3 = 0$$

This can be solved for β_i* if $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \neq 0$.

We could now substitute β_1* and β_2* back into g_3*
and compare this with several possible alternative control modes and
for several classes of disturbance vectors. However at this level
of complexity this sort of work is probably best left for numerical
experimentation on a digital computer.

We can, however, compare this control mode with the case
where the system is reticulated but the interaction between the
reticulated parts is completely neglected. This case differs from
the model just considered in that no second level unit is considered
and hence $\beta_1 = \beta_2 = 1$.

In the model just considered assume that $\overline{\epsilon}^i = \overline{\mu}^i$ with
probability 1 (deterministic disturbance). Let $g_{II}*$ be the optimal
value of the system performance functional for the indirect inter-
vention approach (Model 1). Let $g_{NI}*$ be the optimal value of the
system performance functional for the case of neglected interaction.

Theorem 4.1. $\qquad g_{II}^* \leq g_{NI}^*$

Proof: Let $\quad g_{II}^* = \min\limits_{\beta_1,\beta_2} \; f\,(\beta_1,\ \beta_2,\ \bar{\epsilon}) = f\,(\beta_1^*,\beta_2^*,\ \bar{\epsilon})$

β_1 and $\beta_2 = 1$ are admissable values of the control variables.

But $f\,(1,\ 1,\ \bar{\epsilon}) = g_{NI}^*$. Therefore, since $f\,(\beta_1^*,\ \beta_2^*,\ \bar{\epsilon}) \leq f\,(1,\ 1,\bar{\epsilon})$

$g_{II}^* \leq g_{NI}^*$ Q.E.D.

Next we will prove a theorem which illustrates the difficulty of comparing two control modes on an absolute basis when the input disturbances are stochastic. We can view the processes we have treated as being control problems which arise at regular intervals in time but each interval's outputs and performance functionals are statistically independent of those in preceding intervals.

Theorem 4.2. $\qquad$ Let $\left\{\bar{\epsilon}_i\right\}$ for $i \in I$ (I is a denumerable index set) be a vector-valued independent random process. Let $\bar{y}_i$ be the output variables associated with the sequence of control vectors $\bar{m}_i$. Let $g_i\,(\bar{m}_i,\ \bar{y}_i,\ \bar{\epsilon}_i)$ be a sequence of performance functionals for $i \in I$. Consider two different control modes yielding $\bar{y}_i^1,\ \bar{m}_i^1$ and $\bar{y}_i^2,\ \bar{m}_i^2$ respectively. Assume there exists a sequence $\left\{k\right\}$ for $k \in I$ such that $P\left(g_k\,(\bar{y}_k^1,\ \bar{m}_k^1,\ \bar{\epsilon}_k)\ -\ g_k\,(\bar{y}_k^2,\ \bar{m}_k^2,\ \bar{\epsilon}_k) > 0\right) = P_k > 0$ and further $\sum\limits_{k=1}^{\infty} P_k = \infty$. Then almost surely for an ∞ number of points $i \in I$ the performance of control mode 1 will exceed the performance of control mode 2.

The proof follows directly from the Borel Cantelli Lemmas. See, for example, Doob.[9]

Essentially, then, if it is possible for one control mode to be better than another (for its utility functional to take on a higher value) and further if the sum of the probabilities of this event taken over all time is unbounded then with probability 1 this event will occur not once but an infinite number of times. However, it is still possible for one control mode to be better than a second on the average and still have the second better than the first for an infinite number of points in time. We have stated this theorem to point out the difficulties involved in comparing two control modes on an absolute basis when the system being modeled is subject to probabilistic disturbances.

4.2 Model 2

We next consider a model where the "cost of control" enters the performance functionals as a quadratic term instead of linearly as before.

System 1

$$g_1{}^* = \min_{\overline{m}^1} \left[{}_1\overline{y}^{1^T} B_{11}\, {}_1\overline{y}^1 + \rho_1\, \overline{m}^{1^T} C_{11}\, \overline{m}^1 \right] \qquad (4.8)$$

subject to ${}_1\overline{y}^1 = A_{11}\, \overline{m}^1 + \overline{\epsilon}^1$.

In this model the terms $\overline{\epsilon}^i$ are considered to be deterministic and known. The method of attack for probabilistic disturbances has been outlined in previous models.

System 2

$$g_2* = \min_{\overline{m}^2} \left[{}_2\overline{y}^{2^T} B_{22} \, {}_2\overline{y}^2 + \beta_2 \, \overline{m}^2 \, C_{22} \, \overline{m}^2 \right] \qquad (4.9)$$

subject to $\quad {}_2\overline{y}^2 = A_{22} \, \overline{m}^2 + \overline{\epsilon}^2$.

Again the quantities β_i are "cost" variables manipulated by the second level unit (System 3). From the point of view of the first level systems the β_i are fixed parameters, i.e. they do not anticipate or predict values for these quantities but rather wait for the second level unit to communicate this information to them.

The optimal values for $\overline{m}^1$ and $\overline{m}^2$ are given by

$$\overline{m}^{1*} = - \left[A_{11}^T B_{11} A_{11} + C_{11} \, \beta_1 \right]^{-1} A_{11}^T B_{11} \, \overline{\epsilon}^1 \qquad (4.10)$$

$$\overline{m}^{2*} = - \left[A_{22}^T B_{22} A_{22} + C_{22} \, \beta_2 \right]^{-1} A_{22}^T B_{22} \, \overline{\epsilon}^2$$

Now for the second level unit we have:

$$g_3* = \min_{\beta_1, \beta_2} \left\{ {}_3\overline{y}^T B \, {}_3\overline{y} + \overline{m}*^T C \, \overline{m}* \right\} \qquad (4.11)$$

$$\overline{m}* = \begin{bmatrix} \overline{m}^{1*} \\ \overline{m}^{2*} \end{bmatrix} , \qquad C = \begin{bmatrix} C_{11} & 0 \\ 0 & C_{22} \end{bmatrix} , \qquad B = \begin{bmatrix} B_{11} & 0 \\ 0 & B_{22} \end{bmatrix}$$

$$_3\overline{y} = \begin{bmatrix} {}_3\overline{y}^1 \\ {}_3\overline{y}^2 \end{bmatrix} , \qquad {}_3\overline{y}^1 = A_{11} \, \overline{m}^{1*} + A_{12} \, \overline{m}^{2*} + \overline{\epsilon}^1$$

$$\qquad\qquad\qquad {}_3\overline{y}^2 = A_{21} \, \overline{m}^{1*} + A_{22} \, \overline{m}^{2*} + \overline{\epsilon}^2$$

Again, the second level unit attempts to set β_i in such a way as to compensate for the interaction terms neglected in the models of Systems 1 and 2.

38

Naively we might substitute (4.10) into (4.11) and attempt to carry out the minimization analytically. However, it is here that we begin to run into difficulty; for β_1 and β_2 are contained in the inverse of a fairly complicated matrix and hence enters g_3* in a way which is quite difficult analytically.[t] We will employ the following theorem from Taylor[21] to remedy this state of affairs.

Let $\begin{bmatrix} X \end{bmatrix}$ be the space of all bounded linear operators defined on the Banach space X mapping X into itself.

Theorem 4.3. If $A \in \begin{bmatrix} X \end{bmatrix}$ and $|\lambda| \geq \|A\|$ where $\|A\|$ is the norm of $A \longrightarrow (\bar{\lambda} - A)^{-1}$ exists and is continuous and $(\lambda - A)^{-1} y = \sum_1^\infty \lambda^{-n} A^{n-1} y$ for each y in the range of $\lambda - A$. Further since X is complete $(\lambda - A)^{-1} = \sum_1^\infty \lambda^{-n} A^{n-1}$. (See Appendix (A) for definitions of the terms employed in this section).

Since we have assumed that the vectors $\bar{y} - \bar{\epsilon}$, $\bar{m}$ are defined over Euclidean n-space, the conditions of the theorem regarding A are satisfied.

Now we return to the problem of finding a tractable expression for $\bar{m}_1*$ and $\bar{m}_2*$.

---

[t]In numerical treatments, though, this is not particularly alarming. If all of the numerical values of A, B, C and are specified we have a fairly simple search problem in two variables. However, we would like, if possible, to offer an approximation theory for such problems.

$$\bar{m}_1^* = - \left[A_{11}^T B_{11} A_{11} + C_{11} \; \beta_1 \right]^{-1} A_{11}^T B_{11} \; \bar{\epsilon}^1 \qquad (4.12)$$

$$= - C_{11}^{-1} \left[A_{11}^T B_{11} A_{11} C_{11}^{-1} + \beta_1 \right]^{-1} A_{11}^T B_{11} \; \bar{\epsilon}^1$$

Let $T_1 = - A_{11}^T B_{11} A_{11} C_{11}^{-1}$

$$\bar{m}_1^* = - C_{11}^{-1} \left[\beta_1 - T_1 \right]^{-1} A_{11}^T B_{11} \; \bar{\epsilon}^1$$

Similarly $\bar{m}_2^* = - C_{22}^{-1} \left[\beta_2 - T_2 \right]^{-1} A_{22}^T B_{22} \; \bar{\epsilon}^2 \qquad (4.13)$

where $T_2 = - A_{22}^T B_{22} A_{22} C_{22}^{-1}$.

Now for $|\beta_1| > ||T_1||$ we can invoke theorem 4.3 to express

$$\left[\beta_1 - T_1 \right]^{-1} = \sum_1^\infty \beta_1^{-n} T_1^{n-1} \quad \text{and} \quad \left[\beta_2 - T_2 \right]^{-1} = \sum_1^\infty \beta_2^{-n} T_2^{n-1}.$$

We can now truncate these series at any point depending on the desired degree of approximation and using the resultant approximations for (4.12) and (4.13) we could solve (4.11) analytically. As can be easily verified, the truncation of the series above at the first terms, i.e. $(\beta_1 - T_1)^{-1} \cong \dfrac{1}{\beta_1}$ and $(\beta_2 - T_2)^{-1} \cong \dfrac{I}{\beta_2}$ yield a set

of simultaneous linear equations for the optimal values of β_1 and β_2.

The approximations developed above are valid for

$|\beta_i| > ||T_i||$. On the other hand for small β_i we can approximate

$\left[\beta_i - T_i \right]^{-1}$ by $- T_i^{-1}$ and hence we have a valid approximation for both large and small β_i.

For a qualitative comparison of this model with the case where the interaction between first level units is neglected we

invoke theorem 4.1. The proof must be slightly modified to account for the change in the performance functional from model 1. Hence we can say that no matter what the (deterministic) disturbances $\bar{\epsilon}$ are the indirect intervention approach yields better performance than the one obtained by neglecting the first level interaction.

4.3 Model 3

In the previous model we considered indirect intervention in the case where the "intervention parameter" was inserted in the term which yields the "costs" of manipulated variables. We now wish to insert the "intervention parameter" in the term of the functional which sets the "cost" of $\bar{y}^T B \bar{y}$.

System 1

$$g_1{}^* = \min_{\bar{m}^1} \left[{}_1\bar{y}^{1T} B_{11} \, {}_1\bar{y}^1 \beta_1 + \bar{m}^{1T} C_{11} \, \bar{m}^1 \right] \qquad (4.13)$$

subject to $\quad {}_1\bar{y}^1 = A_{11} \, \bar{m}^1 + \bar{\epsilon}^1$

System 2

$$g_2{}^* = \min_{\bar{m}^2} \left[{}_2\bar{y}^{2T} B_{22} \, {}_2\bar{y}^2 \beta_2 + \bar{m}^{2T} C_{22} \, \bar{m}^2 \right] \qquad (4.14)$$

subject to $\quad {}_2\bar{y}^2 = A_{22} \, \bar{m}^2 + \bar{\epsilon}^2$

The derivative of g_i with respect to $\bar{m}^i$ yields

$$\beta_1 A_{11}^T B_{11} (A_{11} \, \bar{m}^i + \bar{\epsilon}^i) + C_{11} \, \bar{m}^i = 0 \quad \text{and if we let} \quad \beta_i = \frac{1}{\beta_i}$$

we find

$$\overline{m}^{i*} = - \left[A_{ii}^T \; B_{ii} \; A_{ii} + C_{ii} \; \mathscr{A}_i \right]^{-1} \left[A_{ii}^T \; B_{ii} \; \overline{\epsilon}^i \right]. \quad (4.15)$$

If we look back we see that (4.15) has exactly the same form as (4.10) and hence we need go no further. Unless there have been special restrictions placed on the ρ_i , this form of intervention is the same as that in Model 2.

4.4   Indirect Intervention

Let us consider briefly what precisely is involved in what we have called "Indirect Intervention". For a single system or subsystem let us consider a 2n + 1 dimensional "system space", and let the first n axis represent the control vector components $\overline{m}$. We will call this the "control subspace". Along the next n axis are measured the disturbance components $\overline{\epsilon}$. This is the "disturbance space". Along the last axis is measured the "quality" of performance as measured by the performance functional. Therefore it is assumed that if one picks a point in the "control" subspace and a point in the "disturbance" subspace that a particular value of the "quality" component is determined thereby.

We note that in the previous problems

$$\overline{m}^{1*} = \left[m_1 \; (\beta_1, \; \overline{\epsilon}^1) \; , \; m_2 \; (\rho_1, \; \overline{\epsilon}^1) \; \cdots \; m_k \; (\beta_1, \; \overline{\epsilon}^1) \right] \quad \text{and}$$

$$\overline{m}^{2*} = \left[m_{k+1} \; (\beta_2, \; \overline{\epsilon}^2) \; \cdots \; m_n \; (\beta_2, \; \epsilon^2) \right].$$

Now when these vectors were substituted into g_3 we had

$g_3^* = \min\limits_{\beta_1, \beta_2} g_3 (\beta_1, \beta_2, \bar{\mathcal{E}})$. Therefore where we began with a

problem in n controllable variables we now have a problem in

only two manipulated variables. The questions we would like to

answer are: "What is the geometric counterpart of this process?"

and "What did we lose in quality of performance?" The answer to

the last question depends, of course, on the specific system and

the particular disturbance vector $\bar{\mathcal{E}}$. The first question, however,

can be answered as follows: We have introduced a two parameter

family of curves into the "control subspace" and the resultant con-

trol vector must now lie somewhere on this surface. Fig. 4.1

describes a simple case with two m control components and a single

β.

Fig. 4.1
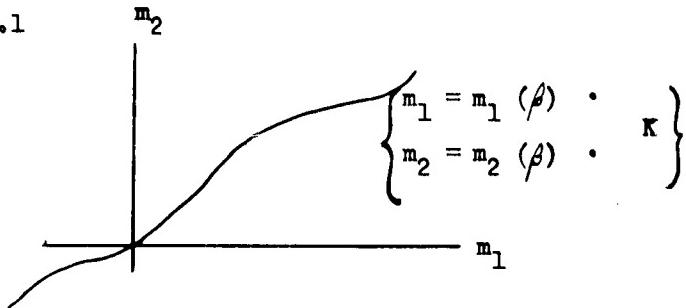
Where originally any point in the m_1, m_2 plane was an admissible

control vector, now only those points lying on the one-parameter

curve K are admissible. Therefore we traded an optimization problem

in two variables for one in a single variable but we reduced the
set of admissible control vectors in the process.

Summary

In this chapter we introduced a new method of second level
"coordination" and we investigated its properties by employing it
in three different models.

Model 1

First Level Units $\quad g_i{}^* = \min\limits_{\bar{m}^i} \quad E\left\{ \bar{y}^{iT} B_{ii} \bar{y}^i + 2\beta_i \bar{1} \cdot \bar{m}^i \right\}$

subject to $\quad \bar{y}^i = A_{ii} \bar{m}^i + \bar{\epsilon}^i \qquad\qquad i = 1, 2$

Second Level Unit $\quad g_3{}^* = \min\limits_{\beta_1, \beta_2} \quad E\left\{ \bar{y}^T B \bar{y} + \bar{c}^T \bar{m} \right\}$

where $\quad \bar{y} = A\bar{m} + \bar{\epsilon}$

Model 2

First Level Units $\quad g_i{}^* = \min\limits_{\bar{m}^i} \quad (\bar{y}^{iT} B_{ii} \bar{y}^i + \beta_i \bar{m}^{iT} C_{ii} \bar{m}^i)$

subject to $\quad \bar{y}^i = A_{ii} \bar{m}^i + \bar{\epsilon}^i \qquad\qquad i = 1, 2$

Second Level Unit $\quad g_3{}^* = \min\limits_{\beta_1, \beta_2} \quad (\bar{y}^T B \bar{y} + \bar{m}^T C \bar{m})$

where $\quad \bar{y} = A\bar{m} + \bar{\epsilon}$

Model 3

First Level Units $\quad g_i^* = \min\limits_{\overline{m}^i} \left(\beta_i \, \overline{y}^{i^T} B_{ii} \, \overline{y}^i + \overline{m}^{i^T} C_{ii} \, \overline{m}^i \right)$

where $\quad \overline{y}^i = A_{ii} \, \overline{m}^i + \overline{\epsilon}^i \qquad\qquad\qquad i = 1, \, 2$

Second Level Unit (Same as Model 2)

Conclusions:

1. For Models 1, 2 and 3 with deterministic distur-
bances, indirect intervention always yields better
performance than that achieved by neglecting the
interaction between the first level systems.

2. Indirect Intervention reduces the controllability
of the original system by reducing the class of
admissible control vectors.

3. It is irrelevant whether the second level unit
fixes the "price" of the control vectors or whether
it sets the penalty "cost" associated with the
outputs. Both models of control are equivalent.

4. It is usually impossible to compare the perfor-
mance of alternate methods of control on an absolute
basis when the system is subject to stochastic
disturbances.

Applications

See Appendix (C).

5. DIRECT INTERVENTION - DYNAMIC CASE

Introduction

We will leave the domain of statics and consider now a class of multi-variable dynamical systems. In particular, our representation of the system dynamics will be in discrete time. The basic dynamic model will be the following:

$$\bar{y}_{t+1} = A \bar{y}_t + \bar{m}_{t+1} + \bar{\epsilon}_{t+1}, \quad \bar{y}_0 = \bar{c} \quad t = 0, 1, \dots \quad (5.1)$$

The quantities $\left\{y_t\right\}$ are n-component vectors of state variables. A is an $n \times n$ non-singular matrix. $\left\{m_t\right\}$ are n-component vectors of manipulated variables. The $\left\{\bar{\epsilon}_t\right\}$ is an n-dimensional vector valued random sequence with known means, i.e. $E(\bar{\epsilon}_t) = \bar{\mu}_t$. We assume

$$E(\bar{\epsilon}_\uparrow^T \bar{\epsilon}_t) = 0 \quad \text{if } t \neq \uparrow \quad \text{and} \quad E(\epsilon_t^T \epsilon_t) < \infty.$$

Mathematical models such as (5.1) may be taken to mean that significant changes in the system take place only at evenly spaced intervals in time. Alternatively, we may interpret (5.1) as observations of a continuous system^t taken at discrete intervals of time. The $\left\{m_t\right\}$ are assumed to be constant in the inter-sampling intervals.

Similarly the presence of random disturbances $\left\{\epsilon_t\right\}$ can be viewed in two different ways. We might assume that

---

^{t}The behavior of linear systems in the inter-sampling period has been explained by control theorists. (11)

$$\bar{y}_{t+1} = A \, \bar{y}_t + \bar{m}_{t+1} \qquad\qquad (5.2)$$

is an exact model of the system dynamics but that certain external influences affect the state variable and are accounted for by the term ϵ_{t+1} in (5.1). Alternatively, we may assume that (5.2) is an inexact model of the real system and hence the term ϵ_{t+1} accounts for the difference between the predicted values of the state variables and those that actually occur. We assume, in both cases, that the $(\bar{\epsilon}\text{'s})$ are observable after the fact, i.e. at time $T = T + 1$, $\bar{\epsilon}_t$, $\bar{\epsilon}_{t-1} \cdots \bar{\epsilon}_1$ are all known.

While it may appear at first glance that a model such as (5.1) is highly restrictive and specialized, certain techniques of substitution can be used to reduce any model of the form

$$\bar{y}_{t+1}^1 = A_1 \, \bar{y}_t^1 + \cdots A_k \, \bar{y}_{t-k+1}^1 + B \, m_{t+1}^1 + C \, \epsilon_{t+1}^1$$

to (5.1) by a suitable relabeling of the variables.

The over-all system optimization problem is:

$$g^* = \max_{\bar{m}_1 \ldots \bar{m}_n} E \left[\sum_{t=1}^{N} \left\{ \bar{y}_t^T \, B \, \bar{y}_t - m_t^T \, C \, m_t \right\} \right] \qquad (5.3)$$

subject to:

$$\bar{y}_{t+1} = A \, \bar{y}_t + \bar{m}_{t+1} + \bar{\epsilon}_{t+1} \qquad\qquad \bar{y}_0 = \bar{c}$$

5.1 Model 1

As in Chapter 3 and 4, we construct the two first-level control problems as follows:

System 1

$$g_1{}^* = \frac{\max}{{}_1\bar{m}_1, {}_1\bar{m}_2 \cdots {}_1\bar{m}_N} \quad E\left\{\sum_{t=1}^{N} {}_1\bar{y}_t^{1^T} B_{11} \, {}_1\bar{y}_t^1 - \bar{m}_t^{1^T} C_{11} \, \bar{m}_t^1\right\} \quad (5.4)$$

${}_1\bar{y}_t$ refers to System 1's model of $\bar{y}_t = \left[y_1(t), y_2(t) \cdots y_k(t)\right]$.

$${}_1\bar{y}_{t+1} = A_{11} \, {}_1\bar{y}_t - + \bar{m}_{t+1}^1 + \bar{\epsilon}_{t+1}^1, \quad {}_1\bar{y}_0 = \bar{c}^1 \quad (5.5)$$

Similarly for System 2:

$$g_2{}^* = \frac{\max}{{}_2\bar{m}_1, {}_2\bar{m}_2 \cdots {}_2\bar{m}_N} \quad E\left\{\sum_{t=1}^{N} {}_2\bar{y}_t^{2^T} B_{22} \, {}_2\bar{y}_t^2 - \bar{m}_t^{2^T} C_{22} \, \bar{m}_t^2\right\} \quad (5.6)$$

$${}_2\bar{y}_{t+1} = A_{22} \, {}_2\bar{y}_t + \bar{m}_{t+1}^2 + \bar{\epsilon}_{t+1}^2, \quad {}_2\bar{y}_0 = \bar{c}^2 \quad (5.7)$$

Now we form a second level system whose function is to compensate for the neglected interaction between the two first level systems.

$$g_3{}^* = \frac{\max}{\bar{\eta}_1, \bar{\eta}_2 \cdots {}_N} \quad E\left\{\sum_{t=1}^{N} {}_3\bar{y}_t^{1^T} B_{11} \, {}_3\bar{y}_t^1 - \bar{\eta}_t^{1^T} C_{11} \, \bar{\eta}_t^1\right\} \quad (5.8)$$

$${}_3\bar{y}_{t+1}^1 = A_{11} \, \bar{y}_t^1 + A_{12} \, \bar{y}_t^1 + \bar{m}_{t+1}^{1*} + \bar{\eta}_{t+1}^1 + \bar{\epsilon}_{t+1}^1$$

and

$$g_4{}^* = \frac{\max}{\bar{\eta}_1, \bar{\eta}_2 \cdots {}_N} \quad E\left\{\sum_{t=1}^{N} {}_4\bar{y}_t^{2^T} B_{22} \, {}_4\bar{y}_t^2 - \bar{\eta}_t^2 C_{22} \, \bar{\eta}_t^2\right\} \quad (5.9)$$

$${}_4\bar{y}_{t+1}^2 = A_{21} \, \bar{y}_t^1 + A_{22} \, \bar{y}_t^2 + \bar{m}_{t+1}^{2*} + \bar{\eta}_{t+1}^2 + \bar{\epsilon}_{t+1}^2$$

It is important to note that in this representation the second level control action is decomposed into the determinations of two functions, $\bar{\eta}_t^{1*}$ and $\bar{\eta}_t^{2*}$. In order to perform this split successfully it is necessary to ignore certain aspects of the system dynamics in order to avoid having the second level unit solve a problem of the same dimensionality as (5.3).

We assume that at time $t+1$ the sample value of the outputs $\bar{y}_t$ are known or measurable by every subsystem without error. We visualize that at time $t+1$ the operation of the decentralized system proceeds as follows:

1. Systems 1 and 2 measure $\bar{y}_t^1$ and $\bar{y}_t^2$ respectively.

2. Systems 1 and 2 optimally determine $\bar{m}_{t+1}^{1*}$ and $\bar{m}_{t+1}^{2*}$ respectively.

3. $\bar{m}_{t+1}^{1*}$ are communicated to the second level unit.

4. Second Level Unit (System 3) measures $\bar{y}_t^i$.

5. System 3 computes $\bar{\eta}_{t+1}^{1*}$ using $\bar{m}_{t+1}^{1*}$ and $\bar{y}_t^i$.

6. System 3 computes $\bar{\eta}_{t+1}^{2*}$ using $\bar{m}_{t+1}^{2*}$ and $\bar{y}_t^i$.

7. All decisions are implemented.

Hence when system 3 determines $\bar{\eta}_{t+1}^1$ it measures $\bar{y}_t^2$ and uses it in the optimization; thus it completely neglects the effect that $\bar{\eta}_{t+1}^1$ will eventually have on $\bar{y}_{t+1}^2$. A similar remark applies to the determination of $\bar{\eta}_{t+1}^2$.

We now proceed to solve the four optimization problems in Dynamic Programming formalism. For System 1 we let

$$S_N^1(\bar{c}^1) = \max_{\bar{m}_1^1} \quad E\left\{ {}_1\bar{y}_1^{1T} B_{11}\, {}_1\bar{y}_1^1 - \bar{m}_1^{1T} C_{11}\, \bar{m}_1^1 \right. \qquad (5.10)$$

$$\left. + S_{N-1}(A_{11}\bar{c}^1 + \bar{m}_1^1 + \bar{\epsilon}_1^1) \right\}$$

where $S_N^1(\bar{c}^1)$ is the maximum expected "return" for the N stage process given the process starts with $\bar{y}_o^1 = \bar{c}^1$. $_1\bar{y}_t^1$ is given by (5.5).

Similarly

$$S_N^2(\bar{c}^2) = \max_{\bar{m}_1^2} \quad E\left\{ {}_2\bar{y}_1^{2T} B_{22}\, {}_2\bar{y}_1^2 - \bar{m}_1^{2T} C_{22}\, \bar{m}_1^2 \right. \qquad (5.11)$$

$$\left. + S_{N-1}^2(A_{22}\bar{c}^2 + \bar{m}_1^2 + \bar{\epsilon}_1^2) \right\}$$

$_2\bar{y}_t^2$ is given by (5.7).

For System 3 we have

$$S_N^3(\bar{c}^1) = \max_{\bar{\eta}_1^1} \quad E\left\{ {}_3\bar{y}_1^{1T} B_{11}\, {}_3\bar{y}_1^1 - \bar{\eta}_1^{1T} C_{11}\, \bar{\eta}_1^1 \right. \qquad (5.12)$$

$$\left. + S_{N-1}^3({}_3\bar{y}_1^1) \right\}$$

$_3\bar{y}_1^1$ is given by (5.8).

50

$$S_N^4 (\bar{c}^2) = \max_{\bar{\eta}_1^2} \quad E \left\{ {}_3\bar{y}_1^{2T} B_{22} \, {}_3\bar{y}_1^2 - \bar{\eta}_1^{2T} C_{22} \, \bar{\eta}_1^2 \right. \tag{5.13}$$

$$\left. + S_{N-1}^4 ({}_4\bar{y}_1^2) \right\}$$

${}_4\bar{y}_1^2$ is given by (5.9).

Note that in each case the actual influence of $\bar{\eta}_{t+1}^1$ on $\bar{y}_{t+1}^2$, and of $\bar{\eta}_{t+1}^2$ on $\bar{y}_{t+1}^1$ is neglected. We also note that if in S_N^3 and S_N^4 we let

$$\bar{\epsilon}_{k+1}^1 + \bar{m}_{k+1}^{1*} + A_{12} \, \bar{y}_k^2 = \bar{\epsilon}_{k+1}^{1T}, \tag{5.14}$$

$$\bar{\epsilon}_{k+1}^2 + \bar{m}_{k+1}^{2*} + A_{21} \, \bar{y}_k^1 = \bar{\epsilon}_{k+1}^{2T},$$

then all four Dynamic Programming problems have the same structure, i.e. the variables in each pair of problems are isomorphic. Hence all are mathematically equivalent to the problem

$$S_N (\bar{c}) = \max_{\bar{m}_1} \quad E \left\{ \bar{y}_1^{T} B \bar{y}_1 - \bar{m}_1^{T} C \bar{m}_1 \right. \tag{5.15}$$

$$\left. + S_{N-1} (A \bar{c} + \bar{m}_1 + \bar{\epsilon}_1) \right\}$$

$\bar{y}_1 = A \bar{c} + \bar{m}_1 + \bar{\epsilon}_1$ or in general $\bar{y}_{t+1} = A \bar{y}_t + \bar{m}_{t+1} + \bar{\epsilon}_{t+1}$

The general solution for $\bar{m}_t$ is of the following form (See Appendix B).

$$\tag{5.16}$$

$$\bar{m}_{N-k}^* = H_{N-k} A \bar{c} + Q_{N-k} \quad \text{where } \bar{c} \text{ is current state}$$

of the system.

$$H_{N-k} = (C - B_{N-k})^{-1} B_{N-k}, \quad B_{N-k} = B + A^T \psi_{N-k+1} A \quad (5.17)$$

$$G_{N-k} = (C - B_{N-k})^{-1} \left[B_{N-k} \bar{\mu}_{N-k} + A^T S_{N-k+1} \right] \quad (5.18)$$

$$\psi_{N-k} = (I + H_{N-k})^T B_{N-k} (I + H_{N-k}) - H_{N-k}^T C H_{N-k} \quad (5.19)$$

$$S_{N-k} = (I + H_{N-k})^T B_{N-k} (\bar{\mu}_{N-k} + G_{N-k}) - H_{N-k}^T C G_{N-k} \quad (5.20)$$

$$+ A^T S_{N-k+1}$$

$$H_N = (C - B)^{-1} B \qquad G_N = (C - B)^{-1} B \bar{\mu}_N \qquad B_N = B \qquad (5.21)$$

$$S_{N+1} = 0 \qquad \bar{\mu}_N = E(\epsilon_N) \qquad (5.22)$$

Therefore

$$\bar{m}_{N-k}^{1*} = H_{N-k}^1 A_{11} \bar{y}_{N-k-1}^1 + G_{N-k}^1 \quad (5.23)$$

$$\bar{m}_{N-k}^{2*} = H_{N-k}^2 A_{22} \bar{y}_{N-k-1}^2 + G_{N-k}^2$$

H_{N-k}^i is identical to H_{N-k} in (5.17) except that $A \longrightarrow A_{ii}$,

$B \longrightarrow B_{ii}$, $C \longrightarrow C_{ii}$, $\bar{\epsilon} \longrightarrow \bar{\epsilon}^i$, $\bar{\mu} \longrightarrow \bar{\mu}^i$. ($\longrightarrow$ means "is replaced by")

Similarly for G_{N-k}^1 and G_{N-k}^2.

For $\bar{n}_{N-k}^{*1}$ and $\bar{n}_{N-k}^{*2}$ we have

$$\bar{n}_{N-k}^{*1} = H_{N-k}^3 \bar{y}_{N-k-1}^1 + G_{N-k}^3 \quad (5.24)$$

$$\bar{n}_{N-k}^{*2} = H_{N-k}^4 \bar{y}_{N-k-1}^2 + G_{N-k}^4$$

$H_N^3 {}_{-k}$ is identical to $H_N^1 {}_{-k}$ except that $\bar{\epsilon}^1 \longrightarrow \bar{\epsilon}^{1^T}$ and

$\bar{\mu}^1 \longrightarrow E\,(\bar{\epsilon}^{1^T})$. Similarly for $H_N^4 {}_{-k}$ and $Q_N^3 {}_{-k}$ and $Q_N^4 {}_{-k}$.

Therefore the true outputs of the system are:

$$\bar{y}_{t+1}^1 = A_{11}\,\bar{y}_t^1 + A_{12}\,\bar{y}_t^2 + \bar{\eta}_{t+1}^{1*} + \bar{\epsilon}_{t+1}^1 + \bar{m}_{t+1}^{1*} \qquad (5.25)$$

$$\bar{y}_{t+1}^2 = A_{21}\,\bar{y}_t^1 + A_{22}\,\bar{t}_t^2 + \bar{m}_{t+1}^{2*} + \bar{\eta}_{t+1}^{2*} + \bar{\epsilon}_{t+1}^2$$

where the m's are given in (5.23) and the η's in (5.24).

Now that we know the specific form of the decision rule
we can conceptually compare the performance of g as given by
(5.3) with, for example, $g_3 + g_4$ as given by (5.8) and (5.9).
These quadratic forms become random variables whose distributions
are determined by the joint distribution of the $\bar{\epsilon}$'s. Middleton[17]
has shown that if the $\bar{\epsilon}$'s are generated by an independent gaussian
random process then any positive definite quadratic form in the
$\bar{\epsilon}$'s is distributed by the Non-Central Chi-Squared distribution.
Approximation techniques[20] have been developed which reduce the
Non-Central Chi-Squared distribution to the Normal distribution and
since the sum and difference of Normal variates is again a Normal
variate we can find the P ($g - g_3 - g_4 > 0$), i.e. that decentral-
ized approach yields a better value of the performance functional
than the "centralized" or "integrated" approach.

A careful examination of equations (5.4) through (5.9)

brings out an interesting conclusion; if we modify the first level

performance functionals appropriately then the second level unit

does not appear to perform a useful function; in this sense, then,

a $1\ell - 2g$ system is equivalent to a $2\ell - 3g$ system. This is due to

the fact that the two functions assigned to the second level with

a view to reduction of the dimensionality, can as well be located

in the first level units.

5.2 Model 2

In the models previously considered we have assumed that

the performance functional was separable, i.e., Block diagonal,

but that the "system" matrix A was not. In this section we will

demonstrate a method which transforms the original problem (5.3) into

one where the "system" matrix is diagonal and hence separable.

Unfortunately, the conditions for separability of the transformed

performance functional are rather restrictive. We note that if

both the performance functional and the A matrix are block diag-

onal, i.e. separable, then the over-all problem can be decomposed

into two or more completely separate problems. In our language this

means that the first level units can work completely separately

and the over-all system performance will not suffer as a result.

We begin by considering (5.3):

$$g^* = \max_{\bar{m}_1 \ldots \bar{m}_N} \ E \sum_{t=1}^{N} \left\{ \bar{y}_t^T B \bar{y}_t - \bar{m}_t^T C \bar{m}_t \right\} \qquad (5.26)$$

where $\quad \bar{y}_{t+1} = A \bar{y}_t + \bar{m}_{t+1} + \bar{\epsilon}_{t+1} \qquad\qquad \bar{y}_o = \bar{c}$

We will assume for this section that the eigen values $\left\{ \lambda_i^a \right\}$ of A

are all distinct and that none are zero. Hence[5] there exists a

transformation S such that $S^T A \ S = \bigwedge_A$ where

$$\bigwedge_A = \begin{bmatrix} \lambda_1^a & 0 & 0 & .. \\ 0 & \lambda_2^a & .... \\ \vdots & 0 & ... \\ 0 & 0 & ... \lambda_n^a \end{bmatrix} \qquad \text{and further } S^T S = I = \begin{bmatrix} 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ 0 & & ..... & 1 \\ \vdots & & & \end{bmatrix}.$$

Therefore from (5.26) we have $\quad S^T \bar{y}_{t+1} = S^T A \ S \ S^T \bar{y}_t + S^T \bar{m}_{t+1}$

$$+ \ S^T \bar{\epsilon}_{t+1} .$$

Now let $\quad \bar{y}_{t+1}^S = S^T y_{t+1}, \quad \bar{m}_{t+1}^S = S^T \bar{m}_{t+1}, \quad$ and

$\bar{\epsilon}_{t+1}^S = S^T \bar{\epsilon}_{t+1} .$ Therefore the performance functional becomes:

$$g^* = \max_{\bar{m}_1^S, \ldots, \bar{m}_N^S} \ E \sum_{t=1}^{N} \left\{ (S \bar{y}_t^S)^T B (S \bar{y}_t^S) - (S \bar{m}_t^S)^T C \quad (5.27) \right.$$

$$\left. (S \bar{m}_t^S) \right\}$$

$$= \max_{\left\{ \bar{m}_i^S \right\}} \ E \sum_{t=1}^{N} \left\{ \bar{y}_t^{S^T} S^T B S \bar{y}_t^S - \bar{m}_t^{S^T} S^T C S \bar{m}_t^S \right\}$$

Now it is known[5] that if A, B, and C are simultaneously

permutable, i.e. if $AB = BA$, $AC = CA$, etc. then

$$S^T B S = \Lambda_B = \begin{bmatrix} b_1 & 0 & 0 & \cdots \\ 0 & \ddots 0 & 0 & \cdots \\ 0 & 0 & \ddots & b_n \\ & \cdots & & \end{bmatrix}, \quad \text{and} \quad S^T C S = \Lambda_C.$$

In this case then we can split the over-all problem not into two
subproblems with several optimization variables in each but into
n single variable problems each of which is completely independent
of the others. Although the aforementioned conditions are rather
restrictive they are well worth examining because of the complete
separability which results.

Summary and Conclusions

5.1 Model 1

A class of system models which possess significant dyna-
mics were formulated and the method of Direct Intervention was dis-
cussed as a method of coordination for the second level unit in a
$2\ell - 3y$ system. The associated analytical work involving the solution
of a Dynamic Programming problem was carried out. The Dynamic Pro-
gramming solution represents a slight extension of the work of
Adorno.[1] The possibility of the comparison of this decentralized
control system with the "integrated" approach were discussed and
the analytic difficulties noted. Approximations were discussed
that could be used for comparative purposes if the system parameters
were known. In conclusion it was noted that this method of inter-
vention leaves much to be desired in that, except possibly as a

device for the division of computational labor, the second level unit serves no useful purpose and that this $2\ell - 3y$ system is equivalent to a $1\ell - 2y$ system.

5.2 Model 2

Sufficient conditions were discussed for the existence of a transformation which completely separates the problem of Model 1. The resultant control optimization equations can be solved one by one independently. In an organizational context the system can be separated into n independent operating divisions without the need for a coordinating unit at the second level.

Applications

See Appendix (C).

6. INDIRECT INTERVENTION - DYNAMIC CASE

Introduction

As we indicated in the introduction, there are two critical problems in optimization. The first is the problem of many state variables, which we have attacked in the preceding chapters. As we saw in the previous chapter, the coupling within the over-all system dynamics may be of such a nature that it is impossible (at least with the models we have been considering) to recouple the system with a second level controller without considering all of the state variables simultaneously. For these cases we demonstrated that a $1\ell - 2\not{q}$ system was equivalent to a $2\ell - 3\not{q}$ system.

We are faced with essentially the same problems in attempting to apply Indirect Intervention from the second level in the dynamic case. Our principal effort in this chapter will be based on an attempt to reduce the search effort required to find the extremizing values of the variables given that it is not feasible to reduce the number of state variables.

The number of state variables alone is not the sole cause of unfeasibility in the computation of optimization problems. The problem of search for the minimum or maximum, as the case may be, is also critical in determining the cost of computation.

Consider a problem in 4 state variables and let us assume we want the solution (numerical) to a problem of the following form.

$$f_N (c_1, c_2, c_3, c_4) = \min_{m_1, m_2, m_3, m_4} \left\{ g (m_1, m_2, m_3, m_4) \quad (6.1) \right.$$

$$\left. + f_{N-1} (T^1 c_1, T^2 c_2, \ldots T^4 c_4) \right\}$$

The c_i are state variables and $T^i = T^i_{(m_1, m_2, m_3, m_4)}$ are trans-

formations which act on the state variables. The particular trans-

formation is determined by $m_1 \ldots m_4$. We assume $f_o (c_1, c_2, c_3,$

$c_4) \equiv 0$. Assume that the c_i $i = 1 \ldots 4$ can take on 10 values

each and $m_1 \ldots m_4$ can take on 10 values each. The 10 values of

each of the m_i determine a grid in the Cartesian product space

$M_1 \times M_2 \times M_3 \times M_4$ which has 10^4 points. If we wish to find the

minimum by searching this grid we must evaluate the right hand side

of (6.1) 10^4 times for every value of c_1, c_2, c_3, c_4 or

$10^4 \cdot 10^4 = 10^8$ times for every value of N. On the fastest digital

computer this would take on the order of 10^5 seconds/N. We will

show that by using Indirect Intervention we can reduce the number

of necessary operations by a factor of 10^2 or 10^3, provided we are

willing to risk a decrease in the quality of system performance in

order to obtain a solution to the computational problem. While we

realize that more efficient search procedures exist than the complete

examination of the entire grid it should be realized that we are

allowing only 10 sec for the location of the minimum of a function

in 4 space in the time comparisons above. In addition, in the

case of 1 or 2 manipulated variables where the objective function

in concave, Fibonacci[4] search techniques can be used which are

extremely efficient.

6.1 Model 1

We begin by considering a problem where, as before, the over-all system dynamics are given by:

$$\bar{y}_{t+1} = A\,\bar{y}_t + \bar{m}_{t+1} + \bar{\varepsilon}_{t+1} \tag{6.1}$$

$\bar{y}_0 = \bar{c}$ where the $\{\bar{\varepsilon}_i\}$ have the same properties as outlined in Chapter 5.

In this section we will not concern ourselves with the utility functional of the second level system but instead we will concentrate on examining the relationship between the first and second levels.

System 1 is given the following problem:

$$g_1^* = \min_{\bar{m}_1,\ldots\bar{m}_N} E\left\{\sum_{t=1}^{N} {}_1\bar{y}_t^{1\,T} B_{11}\,{}_1\bar{y}_t^1 + 2\bar{d}_t^{1\,T}\bar{m}_t^1\right\} \tag{6.2}$$

where ${}_1\bar{y}_{t+1}^1 = A_{11}\,{}_1\bar{y}_t^1 + \bar{m}_{t+1}^1 + \bar{\varepsilon}_{t+1}^1$ $\bar{y}_0 = \bar{c}^1$ $t = 0,\ldots N-1$

The vector sequence $\bar{d}_i^1$ can be viewed as a sequence of "price" or "accounting" vectors which are set by the second level and which specify to the first level unit the "cost" of a unit of each of the components of the control vector $\bar{m}_i$. An expression similar to (6.2) holds for System 2. The dimensionality of $\bar{d}_i^1$ is assumed to be the same as $\bar{m}_i^1$; similarly for $\bar{d}_i^2$ and $\bar{m}_i^2$.

The over-all operation of the system described here proceeds
as follows:

1. Information
   a) System 3 (Second Level) knows $E(\bar{\epsilon}_1) = \bar{\mu}_1$.
   b) System 3 knows A.
   c) System 3 knows complete structure of first level systems, i.e. (6.2) and its counterpart for System 2.
   d) First level units know $\bar{\mu}_1^1$ and $\bar{\mu}_1^2$ respectively.
   e) First level units know (6.2) and counterpart respectively.
2. Operation
   a) First level units optimize with arbitrary $\bar{d}_t^i$.
   b) First level units communicate to System 3 $\bar{m}_t^{i*}(\bar{d}_t^i)$.
   c) Second level unit sets $\bar{d}_t^i$ so as to optimize over-all system performance.

Now we will actually solve the problem specified by 2a)
because of the rather interesting nature of the solution.

Due to the fact that the structure of the solution is
the same for both first level units we will drop the complicated
sub and superscripting used in (6.2).

We let $S_k(\bar{c})$ be the minimum expected value of the per-
formance functional given that we have k decisions yet to make
and given that we follow an optimal policy from the current state $\bar{c}$.

$$S_N(\bar{c}) = \min_{m_1} E\left\{(A\bar{c} + \bar{m}_1 + \bar{\epsilon}_1)^T B(A\bar{c} + \bar{m}_1 + \bar{\epsilon}_1) \right. \tag{6.3}$$

$$\left. + 2\bar{d}_1^T \bar{m}_1 + S_{N-1}(A\bar{c} + \bar{m}_1 + \bar{\epsilon}_1)\right\}$$

Similarly

$$S_1(\bar{c}) = \min_{\bar{m}_N} E\left\{(A\,\bar{c} + \bar{m}_N + \bar{\varepsilon}_N)^T B (A\,\bar{c} + \bar{m}_N + \bar{\varepsilon}_N) \right. \tag{6.4}$$

$$\left. + 2\,\bar{d}_N^T\,\bar{m}_N\right\}$$

$$\bar{m}_N^* = -B^{-1}\,\bar{d}_N - (A\,\bar{c} + \bar{\mu}_N) \tag{6.5}$$

Extending this process recursively (See Appendix (B)) we obtain the following expression for the sequence of optimal control vector:

$$\bar{m}_{N-k}^* = B^{-1}(A^T\,\bar{d}_{N-k+1} - \bar{d}_{N-k}) - (A\,\bar{c} + \bar{\mu}_{N-k}) \tag{6.6}$$

$$k = 1, \ldots N - 1$$

This has some interesting features. First, since there are as many components in $\left\{\bar{d}_k\right\}$ as in $\left\{\bar{m}_k\right\}$, this implies that this method of intervention does not reduce the controllability of the original system. Therefore, if the system manager could have achieved a given performance with the integrated approach, then the second level unit can achieve the same performance with this mode of intervention.

Second, since the values of $\bar{d}_{N-k}$ and $\bar{d}_{N-k+1}$ are set by the second level unit, they must be communicated to the first level unit prior to the time it implements $\bar{m}_{N-k}^*$. However, in order that the second level unit be able to set $\bar{d}_{N-k+1}$, it customarily requires knowledge of the state of the system <u>after</u> stage N-k. Therefore

if the second level unit does not predict the optimal value of

$d_{N - k + 1}$ and communicate it to the first level system, then the

first level system is forced to do this prediction.

If we interpret this result in an organization-like context

we see (at least for this class of dynamic problems) that the planning

function for the manager, i.e. prediction, is a natural outgrowth

of the organizational structure and the environment.

6.2 Model 2

In this section we consider a slightly different form of

performance functional for the first level systems. We will see that

the qualitative performance of the organizational structure changes

abruptly from that of Model 1 contrary to what we might expect.

System 1

$$g_1^* = \min_{\bar{m}_1^1, \ldots \bar{m}_N^1} E \sum_{t=1}^{N} (_1\bar{y}_t^{1^T} B_{11} \, _1\bar{y}_t^1 + \beta_t^1 \, \bar{m}_t^{1^T} C_{11} \, \bar{m}_t^1) \qquad (6.7)$$

$\left\{ \beta_t^1 \right\}$ are scalars set by the second level.

$_1\bar{y}_{t+1}^1$ is given by (6.2).

System 2

$$g_2^* = \min_{\bar{m}_1^2, \ldots \bar{m}_N^2} E \sum_{t=1}^{N} (_2\bar{y}_t^{2^T} B_{22} \, _2\bar{y}_t^2 + \beta_t^2 \, \bar{m}_t^{2^T} C_{22} \, \bar{m}_t^2) \qquad (6.8)$$

$$_2\bar{y}_{t+1}^2 = A_{22} \, _2\bar{y}_t^2 + \bar{m}_{t+1}^2 + \bar{\epsilon}_{t+1}^2$$

Again the optimal control vectors for the first level units will have the same structure and hence we drop the unnecessary subscripts before we begin to derive a solution.

As in Model 1:

$$\qquad (6.9)$$

$$S_N (\bar{c}) = \min_{\bar{m}_1} \quad E \left\{ \bar{y}_1^T B \bar{y}_1 + \beta_1 \bar{m}_1^T C \bar{m}_1 + S_{N-1} (A \bar{c} + \bar{m}_1 + \bar{\epsilon}_1) \right\}$$

and $S_1 (\bar{c}) = \min_{\bar{m}_N} \quad E \left\{ \bar{y}_N^T B \bar{y}_N + \beta_N \bar{m}_N^T C \bar{m}_N \right\}$

$$\therefore \bar{m}_N^* = - \left[B + \beta_N C \right]^{-1} B (A \bar{c} + \bar{\mu}_N)$$

$$\qquad (6.10)$$

$$\bar{y}_N = \left[I - (B + \beta_N C)^{-1} B \right] A c - \left[B + \beta_N C \right]^{-1} B \bar{\mu}_N + \bar{\epsilon}_N$$

$$= a_N \bar{c} + \bar{\delta}_N + \bar{\epsilon}_N$$

Hence by substitution we see that $S_1(\bar{c})$ is of the form

$$S_1(\bar{c}) = \bar{c}^T A_N \bar{c} + b_N^T \bar{c} + d_N \qquad (6.11)$$

where A_N is a matrix, b_N is a vector and d_N is a scalar.

Hence

$$S_2(\bar{c}) = \min_{\bar{m}_{N-1}} E \left\{ (A\bar{c} + \bar{m}_{N-1} + \bar{\epsilon}_{N-1})^T B \qquad (6.12) \right.$$

$$(A\bar{c} + \bar{m}_{N-1} + \bar{\epsilon}_{N-1}) + \beta_{N-1} \bar{m}_{N-1}^T C \bar{m}_{N-1}$$

$$+ (A\bar{c} + \bar{m}_{N-1} + \bar{\epsilon}_{N-1})^T A_N (A\bar{c} + \bar{m}_{N-1} + \bar{\epsilon}_{N-1})$$

$$\left. + b_N^T (A\bar{c} + \bar{m}_{N-1} + \bar{\epsilon}_{N-1}) + d_N \right\}$$

Taking the expectation and differentiating, we have

$$\bar{m}_{N-1}^* = (B + \beta_{N-1} C + A_N)^{-1} \left[B(A\bar{c} + \bar{\mu}_{N-1}) \qquad (6.13) \right.$$

$$\left. + A_N (A\bar{c} + \bar{\mu}_{N-1}) + \frac{b_N}{2} \right]$$

If we examine (6.9) and (6.13) carefully we can detect some interesting aspects of the structure of these solutions. We see that β_N enters into $\bar{m}_N^*$ and $\bar{m}_{N-1}^*$ contains β_{N-1}. However $\bar{m}_{N-1}^*$ also contains β_N in the terms A_N and b_N. Similarly we see that $\bar{m}_{N-2}^*$ will contain β_N, β_{N-1}, and β_{N-2}.

In other words, the second level system must predict the "price" vectors from the current point in time to the end of the

decision horizon. This might look at first glance as though it would overload the second unit to the point where it would make the model entirely useless. Actually, however, it opens up possibilities for a large number of interesting alternatives. One of these is an adaptive prediction scheme where the second level system gives initial predictions which allows the first level units to find the first few control vectors and as the process proceeds the second level unit would update its predictions based on further information.

Analytically this model is quite cumbersome and hence we will not pursue the manipulation further.

We will now explore a class of models which reduce the search problem associated with optimization. Model 1 does, however, show us that we can achieve the true optimum in performance with indirect intervention if we are willing to use enough components in the $\bar{d}_i$ vectors.

6.3 Model 3

We now consider a model where the second level sets the internal prices $\bar{d}^1$ only once during the N stage process.

System 1

$$g_1^* = \min_{\left\{\bar{m}_t^1\right\}} E\left\{\sum_{t=1}^N {}_1\bar{y}_t^{1^T} B_{11} {}_1\bar{y}_t^1 + 2\bar{d}^1 \bar{m}_t^1\right\} \qquad (6.14)$$

where $\quad {}_1\bar{y}_{t+1}^1 = A_{11} {}_1\bar{y}_t^1 + \bar{m}_{t+1}^1 + \bar{\epsilon}_{t+1}^1 \qquad \bar{y}_0^1 = \bar{c}^1$

Similarly for System 2

$$g_2^* = \min_{\overline{m}_t^2} \quad E\left\{ \sum_{t=1}^{N} \; {}_2\overline{y}_t^{2^T} B_{22} \; {}_2\overline{y}_t^2 + 2\overline{d}^2 \; \overline{m}_t^2 \right\} \tag{6.15}$$

$${}_2\overline{y}_{t+1}^2 = A_{22} \; {}_2\overline{y}_t^2 + \overline{m}_{t+1}^2 + \overline{\epsilon}_{t+1}^2 \qquad \overline{y}_0^2 = \overline{c}^2$$

For System 3:

$$g_3^* = \min_{\overline{d}^1, \overline{d}^2} \quad E\left\{ \sum_{t=1}^{N} \; {}_3\overline{y}_t^{1^T} B_{11} \; {}_3\overline{y}_t^1 + {}_3\overline{y}_t^{2^T} B_{22} \; {}_3\overline{y}_t^2 \right\} \tag{6.16}$$

$${}_3\overline{y}_{t+1}^1 = A_{11} \; {}_3\overline{y}_t^1 + A_{12} \; {}_3\overline{y}_t^2 + \overline{m}_{t+1}^{1^*} + \overline{\epsilon}_{t+1}^1$$

$${}_3\overline{y}_{t+1}^2 = A_{21} \; {}_3\overline{y}_t^1 + A_{22} \; {}_3\overline{y}_t^2 + \overline{m}_{t+1}^{2^*} + \overline{\epsilon}_{t+1}^2$$

From (6.6) upon deleting the subscripts we have

$$\overline{m}_t^{i^*} = B_{ii}^{-1} \; (A_{ii}^T - I) \; \overline{d}^i - A_{ii} \; \overline{y}_{t-1}^i - \overline{\mu}_t^i \tag{6.17}$$

This can now be substituted into (6.16) and the optimization carried out.

It is easily seen that the model just considered significantly reduces the computational problems for the system. First of all, each of the first level systems is working with a model with approximately one-half of the original variables. The second level unit is required to carry out only one optimization in the N periods, instead of N as in Model 1.

It may be of interest for the purposes of exposition to interpret Models 1 and 2 in an organizational context. In Model 1

each worker (first level unit) is paid a fluctuating salary which may
be different from day to day. The manager manipulates his salary in
order to get the optimal amount of work from him. In the second
model each worker's salary is set once and for all (for the N periods),
but again the manager would like to set the salary structure in a
way which will yield the best system performance. This second model
might well form the basis for a model of "management by exception",
where the manager reviews the salary structure from time to time and
adjusts the salary structure based on a review of the over-all system
performance.

Summary and Conclusions

In this chapter we examined three different models with
an eye to reducing the dimensionality of search in the optimization.

Model 1. In this model we found that if we were willing
to include enough components in the "intervention vector" $\bar{d}_i$ that
we could achieve the same result with decentralized control as with
the integrated approach. We also found that for this class of
models the planning function arose naturally as a result of the or-
ganizational structure.

Model 2. Here we investigated a model which was a slightly
modified version of Model 1. Here it was found that the qualitative
characteristics of the system operation changed abruptly. The
second level unit was required to predict the "prices" for the first

level systems over the entire decision horizon instead of just one
period ahead as in Model 1.

Model 3. Here the second level set one "price" vector for
each of the first level units and then left them alone. This pro-
cedure significantly reduces the dimensionality of the search problem.

No models were formulated for the derivation of the optimal
internal "prices" by the second level units. This is an entirely
straight-forward process and would proceed in a manner similar to
the formulation of the models for the first level units. It should
be pointed out again that the second level unit will have to include
all of the original state variables in its optimization, but that the
search will be over a set of significantly reduced dimensionality
(in Models 2 and 3).

7. INDIRECT MODEL - DIRECT INTERVENTION

Introduction

The models considered previously have had a common property: They all required that the second level unit have complete knowledge of the structure of all the systems below it and in addition that the first level units communicate their decisions to the second level before they were implemented. Clearly this places a great "information load" on the managerial unit. It is the purpose of this chapter to place a correspondence between solving a particular class of non-linear optimization problems and the operation of a decentralized organization. The algorithm developed for the optimization problem will make possible the synthesis of an organizational structure which significantly reduces the problems of "information overload".

7.1 The Programming Problem

We will consider problems of the following type.

$$\max_{\bar{m}^1,\bar{m}^2} \quad g_1(\bar{m}^1) + g_2(\bar{m}^2) \qquad \text{subject to} \qquad (7.1)$$

$$A_1(\bar{m}^1) \leq a_1 \qquad\qquad a_i \geq 0$$
$$\vdots$$
$$A_j(\bar{m}^1) \leq a_j$$

$$A_{j+1}(\bar{m}^1) + B_{\ell+1}(\bar{m}^2) \leq K \qquad\qquad K \geq 0$$

$$B_1(\bar{m}^2) \leq b_1$$
$$\vdots \qquad\qquad\qquad b_i \geq 0$$
$$B_\ell(\bar{m}^2) \leq b_\ell$$

where $\bar{m}^1 = \left[m_1, \ldots m_k \right]$, $\bar{m}^2 = \left[m_{k+1}, \ldots m_N \right]$ and

g_i, $-A_i$, $-B_i$ are all assumed to be positive concave functions of their respective arguments. In this section we will develop a "Decomposition" principle for the solution of (7.1) so that the problem can be broken into subproblems and the resultant solution of these subproblems will then be "coordinated" to yield a solution to (7.1).

The subproblems are:

$$\max_{\bar{m}^1} \quad g_1 (\bar{m}^1) + \lambda_1 (S_1 - A_{j+1} (\bar{m}^1)) \qquad (7.2)$$

subject to $A_1 (\bar{m}^1) \leq a_1$

$A_2 (\bar{m}^1) \leq a_2$

$A_j (\bar{m}^1) \leq a_j$

$$\max_{\bar{m}^2} \quad g_2 (\bar{m}^2) + \lambda_2 (S_2 - B_{\ell+1} (\bar{m}^2)). \qquad (7.3)$$

subject to $B_1 (\bar{m}^2) \leq b_1$

$\vdots$

$B_\ell (\bar{m}^2) \leq b_\ell$

The terms $\lambda_1 (S_1 - A_{j+1} (\bar{m}^1))$ and $\lambda_2 (S_2 - B_{\ell+1} (\bar{m}^2))$

require some explanation. The factors λ_1 and λ_2 are Lagrange multipliers. S_1 and S_2 are variables such that $0 \leq S_1 \leq K$, $0 \leq S_2 \leq K$. By a "formal solution" to problems (7.2) and (7.3) we mean $\bar{m}^{1*} (S_1)$, $\lambda_1^* (S_1)$ and $\bar{m}^{2*} (S_2)$, $\lambda_2^* (S_2)$ respectively,

where $\overline{m}^{-1*}(S_1)$ is the vector of optimizing values of $\overline{m}^{-1}$ as a function of the value of S_1 and $\lambda_1^*(S_1)$ is the value of the Lagrange multiplier at the optimal point as a function of S_1. Similar remarks hold for $\overline{m}^{-2*}(S_2)$ and $\lambda_2^*(S_2)$.

We now require the following lemma on Lagrange multipliers.

Lemma 7-1. Given the problem

$$\max_{\overline{m}^{-1},\overline{m}^{-2}} \quad g(\overline{m}^{-1},\overline{m}^{-2}) + \lambda_1(S_1 - A(\overline{m}^{-1})) + \lambda_2(S_2 - B(\overline{m}^{-2})) \quad (7.4)$$

where, as above, $0 \leq S_i \leq K$ and further g, $-A$ and $-B$ are concave continuously differentiable functions defined over the positive orthant of E^n (Euclidean n space).

Define $g^*(S_1,S_2) \equiv g(\overline{m}^{-1*}(S_1), \overline{m}^{-2*}(S_2))$

$$\therefore \quad \frac{\partial g^*}{\partial S_1} = \sum_{j=1}^{K}\left(\frac{\partial g}{\partial m_j}\bigg|_{\substack{S_1 \\ m_j^*}} \frac{\partial m_j^*}{\partial S_1} + \lambda_1^* - \lambda_1^* \frac{\partial A}{\partial m_j}\bigg| \frac{\partial m_j^*}{\partial S_1} \right)$$

$$= \sum_{j=1}^{n} \frac{\partial g}{\partial m_j^*} \frac{\partial m_j}{\partial S_1} + \lambda_1^* - \lambda_1^* \frac{\partial A}{\partial m_j^*} \frac{\partial m_j^*}{\partial S_1}$$

$$= \lambda_1^* + \sum_{j=1}^{n}\left(\frac{\partial g}{\partial m_j^*} - \lambda_1^* \frac{\partial A}{\partial m_j^*} \right) \frac{\partial m_j^*}{\partial S_1} \quad (7.5)$$

Then whenever $\dfrac{\partial m_j}{\partial S_1}$ is finite, we have $\dfrac{\partial g^*}{\partial s_1} = \lambda_1^*$ because

$\dfrac{\partial g}{\partial m_j^*} - \lambda_1^* \dfrac{\partial A}{\partial m_j^*} = 0$ is a necessary condition for the existence of

an extremum (See, for example, Karlin[13]). Similarly $\dfrac{\partial g^*}{\partial s_2} = \lambda_2^*$

whenever $\dfrac{\partial m_j^*}{\partial s_2}$ is finite. Further $\lambda_1^* \geq 0$, i.e. g^*, is a monotone

function of S_1 .

Now referring to (7.1) we would like to replace

$g = g_1 (\overline{m}^1) + g_2 (\overline{m}^2)$ with $G = G_1 (S_1) + G_2 (S_2)$ where

$G_i (S_i) = g_i (\overline{m}^{i*} (S_i))$.

Hence we require

Theorem 7-1

If $S_1 = 0 \longrightarrow \overline{m}^1 \equiv 0$

$S_2 = 0 \longrightarrow \overline{m}^2 \equiv 0$

then:

$$G_1 (S_1) = \int_0^{S_1} \lambda_1 (\gamma) \, d\gamma + g_1 (0) \qquad \text{and} \qquad (7.6)$$

$$G_2 (S_2) = \int_0^{S_2} \lambda_2 (\gamma) \, d\gamma + g_2 (0) \qquad (7.7)$$

Proof: From Lemma 7-1 we know that

$$\frac{\partial G}{\partial S_1} = \lambda_1 (S_1) \qquad \text{and} \qquad (7.8)$$

$$\frac{\partial G}{\partial S_2} = \lambda_2 (S_2) \qquad (7.9)$$

and further we know by inspection of (7.1) that $G = G_1 + G_2$.

Integrating (7.8) we find $G = \int_0^{S_1} \lambda_1 (\gamma) \, d\gamma + H (S_2)$ where

$H (S_2)$ is an arbitrary function of S_2 . But (7.9) tells us that

$H (S_2) = \int_0^{S_2} \lambda_2 (\gamma) \, d\gamma + C$ where C is an arbitrary constant.

But the hypothesis of the theorem allows us to evaluate C as

$g_1 (\bar{0}) + g_2 (\bar{0})$.

Problem (7.1) can now be reformulated as

$$\max_{S_1, S_2} \quad G_1 (S_1) + G_2 (S_2) \qquad (7.10)$$

subject to $0 \leq S_1 + S_2 \leq K, \quad S_i \geq 0$

There are several sifficulties which may have already come
to the reader's attention. The first question which arises is,
"In what sense is (7.10) equivalent to (7.1)?" We might know the
optimal values of the S_i and not know $\bar{m}^{i*}$. This is due to the
fact that we eliminated quite a lot of information in arriving at
(7.10). Secondly, there is the problem of non-uniqueness. We wrote
$\bar{m}^{i*} (S_i)$ as though it were a specific value, but actually it is a
set of values for there may be many vectors $\bar{m}^i$ which optimize (7.1)

for any specific allocation of the S_i. Third, the analytic difficulties in arriving at (7.10) will generally become insuperable for most problems. In the next section we propose an operational procedure interpreted in the context of the theory of the firm which overcomes these difficulties.

7.2 Interpretation and Operation

We assume that (7.1) arises in the following context. The performance functional is the profit function for a firm with two operating divisions, i.e. g_1 is the profit function for division 1 and g_2 is the profit function for division 2. The constraints A_i $1 \le i \le j$ are given by the technology of division 1. B_i $1 \le i \le \ell$ are similarly defined for division 2. The constraints $A_{j+1} + B_{\ell+1} \le K$ is related to the operation of the firm as a whole. For example, this constraint might be the expression of the fact that the total quantity of operating capital available is equal to K. One approach to the management of this firm is to centralize the operation and give one organizational unit the task of solving (7.1) and setting the production schedules for the operating divisions. This corresponds in mathematical formalism to a direct solution of (7.1). A second approach to the management problem is to set up a planning group for each operating division and a central planning unit to coordinate the activities of these divisional planning groups.

Hence in the delegation of responsibility we give division
1 the task specified by (7.2) and division 2 the task of solving
(7.3). The central planning unit's (unit 3) job is to allocate the
total constraint K between the two operating divisions in such a
way as to maximize the total profit for the firm, i.e. Maximize
$g_1 + g_2$. Now, if we examine (7.10) carefully we see that this is
the mathematical formulation of the central planning unit's problem.
Hence the management of the firm's operations proceeds as follows:

Step 1. The two first level units solve (7.2)
and (7.3) for arbitrary $0 \leq S_i \leq K$ and the
Lagrange multipliers are communicated to the
central planning unit.

Step 2. The central planning unit then uses the
Lagrange multipliers to construct and solve
(7.10). The optimal S's are communicated to
the first level unit.

Step 3. The solutions to (7.2) and (7.3) cor-
responding to the optimal S's are implemented.

This procedure, by construction, has the property that it
yields the solution to (7.1). Hence in the spirit of the previous
chapters we would say that the performance of the dencentralized
system is as good as the integrated (centralized) system.

Perhaps at this point a brief specific example will aid
the reader's intuition. Corresponding to (7.1), consider

$$\max \quad g = m_1 m_2 + m_3^2 \tag{7.1}[1]$$

$$\text{subject to} \quad m_1^2 + m_2 + m_3^4 \le 10$$

$$m_3^2 + m_3 \le 2$$

$$m_i \ge 0$$

We will separate the problem between m_2 and m_3 and hence we see that the problem satisfies all of the necessary[*] conditions for separability (see 7.5).

$$\max \quad g_1 = m_1 m_2 + \lambda_1 \ (S_1 - m_1^2 - m_2) \tag{7.2}[1]$$

$$m_i \ge 0 \qquad i = 1, 2$$

$$\max \quad g_2 = m_3^2 + \lambda_2 \ (S_2 - m_3^4) \tag{7.3}[1]$$

$$\text{subject to} \quad m_3^2 + m_3 \le 2 \quad \text{i.e.} \quad m_3 \le 1$$

$$m_3 \ge 0$$

From $(7.2)[1]$ and $(7.3)[1]$ $\quad \lambda_1^* = \sqrt{\dfrac{S_1}{3}}, \quad \lambda_2^* = \dfrac{1}{2} S_2^{-1/2}, \quad S_2 \le 1$

$(7.10)[1]$ becomes

$$\max \quad g_3 = \frac{2}{3\sqrt{3}} S_1^{3/2} + S_2^{1/2}$$

$$\text{subject to} \quad 0 \le S_1 + S_2 \le 10$$

$$S_2 \le 1 \qquad S_i \ge 0$$

---

[*](The objective function g is not concave, however in this case we will be able to handle the entire problem analytically and we will not need to employ the Gradient Method.)

The solution to $(7.10)^1$ is $S_1^* = 10$, $S_2^* = 0$ which yields $m_3^* = 0$, $m_1^* = \sqrt{\dfrac{10}{3}}$, $m_2^* = \dfrac{20}{3}$ and we note that m_1^* substituted into $(7.1)^1$ is, in fact, equal to g_3^*.

This procedure will become increasingly complicated as the complexity of (7.1) increases and hence we require a numerical procedure. We assume that problem (7.2) and (7.3) can be solved numerically via the use of standard linear or non-linear programming techniques.

We address ourselves here to the problem of the second level unit. This unit receives the numerical values of the respective Lagrange multipliers and from this information alone it must decide on a method for changing the values of the relative allocations of the S_i. For the case where G_1 and G_2 are concave functions of their respective arguments the "Gradient Method" furnishes us with just such a procedure. For a complete description of this method and its stability and convergence properties we refer the reader to the literature.[2] Operationally the use of the "Gradient Method" would proceed as follows:

Step 1. The central planning unit specifies initial allocations of S_1 and S_2. We will denote these by $S_1^{(1)}$ and $S_2^{(1)}$ and $S_1^{(i)} + S_2^{(i)} = K$ for $i = 1, 2 \ldots$

Step 2. The first level units solve (7.2) and (7.3) with S_1 replaced by $S_1^{(1)}$, S_2 replaced by $S_2^{(1)}$.

Step 3. The numerical values of the Lagrange multipliers $\lambda_1^{(1)}$ and $\lambda_2^{(1)}$ are transmitted to the central planning unit.

Step 4. If $\lambda_1^{(1)} > \lambda_2^{(1)}$ then the value of

$$S_1^{(2)} = S_1^{(1)} + \lambda_1^{(1)} - \lambda_2^{(1)}, \; S_2^{(2)} = S_2^{(1)} - \lambda_1^{(1)} + \lambda_2^{(1)}$$

This process proceeds until $\lambda_1^{(i)} = \lambda_2^{(i)}$, at which time the optimum

has been reached at an interior point or it proceeds until

A: $S_1^{(i)} = K$, $S_2^{(i)} = 0$ and $\lambda_1^{(i)} \geq \lambda_2^{(i)}$ or B: $S_1^{(i)} = 0$, $S_2^{(i)} = K$

and $\lambda_2^{(i)} \geq \lambda_1^{(i)}$, i.e. the extremum occurs at an end point. To see

more clearly where the term $\lambda_1 - \lambda_2$ comes from in Step 4 we notice

that $G(S_1, S_2) = G_1(S_1) + G_2(S_2) = G_1(S_1) + G_2(K-S_1)$

hence $\dfrac{\partial G}{\partial S_1} = \dfrac{\partial G_1}{\partial S_1} + \dfrac{\partial G_2}{\partial S_2} \dfrac{\partial S_2}{\partial S_1} = \dfrac{\partial G_1}{\partial S_1} - \dfrac{\partial G_2}{\partial S_2} = \lambda_1 - \lambda_2$

Therefore the changes in the S_i are made proportional to the derivative of the total profit function, hence the name "Gradient Method."

We note in passing that we can accommodate some relaxation of the conditions posed in the formulation of (7.1). First we can

accommodate more than one coupling restriction. Just how many more
is not an easy question to answer, since it depends on the computa-
tional complexity of the specific problem. Second we could with
ease extend this to more than two divisions. Third we could accom-
modate any utility function on the part of the central planning
unit of the form $H (G_1 (\overline{m}_2^{-1}), G_2 (\overline{m}^2))$, if the resultant computa-
tional problem of maximizing H is feasible. Note that H need
not be concave or convex. This again brings up the interesting
possibility that the central planning unit may give the separate
divisions erroneous profit equations to work with to expedite the
operation of the entire firm. The actual feasibility of this pro-
cedure has to be evaluated in the particular context of the specific
problem. Fourth we can at least open the door to and formulate the
problem of more general structures than have previously been con-
sidered in this paper, i.e. the problem of multiple-level structures,
e.g. $m \leq 3$. In our problem K was considered to be fixed by
harsh reality. In fact, however, this could have been set by a
higher level than the central planning unit. The higher level unit
then has the responsibility for coordinating 2 or more (3-unit)
systems of the type we consider here. In the next section we will
formulate this problem in detail.

7.3 Multiple-Level Structures

Consider the following problem:

$$\max_{\overline{m}^1, \overline{m}^2, \overline{m}^3, \overline{m}^4} g_1'(\overline{m}^1) + g_2(\overline{m}^2) + g_3(\overline{m}^3) + g_4(\overline{m}^4) \qquad (7.11)$$

subject to

$$A_{\alpha+1}(\overline{m}^1) + B_{\beta+1}(\overline{m}^2) + C_{\gamma+1}(\overline{m}^3) + D_{\delta+1}(\overline{m}^4) \leq \underline{K}$$

$$A_1(\overline{m}^1) \leq K_1$$
$$\vdots$$
$$A_\alpha(\overline{m}^1) \leq K_\alpha$$

$$B_1(\overline{m}^2) \leq L_1$$
$$\vdots$$
$$B_\beta(\overline{m}^2) \leq L_\beta$$

$$C_1(\overline{m}^3) \leq N_1$$
$$\vdots$$
$$C_\gamma(\overline{m}^3) \leq N_\gamma$$

$$D_1(\overline{m}^4) \leq P_1$$
$$\vdots$$
$$D_\delta(\overline{m}^4) \leq P_\delta$$

We will consider the correspondence between the solution of (7.11) and the problems of decentralized control which might arise in an organizational structure of the following topology.
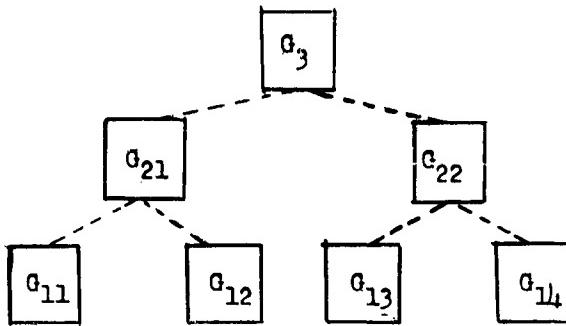
G_3 represents the 3d level goal-seeking unit or system manager. G_{21} and G_{22} are the second level units (middle management). G_{1i} are the first level units. In the language of the General Systems Theory, we have a $3\ell - 7g$ system.

We entrust G_3 with the task of solving (7.11), but we only allow it two manipulated variables. Its problem is to split up the constraint involving the quantity $\underline{K}$ and to allocate the total quantity $\underline{K}$ to G_{21} and G_{22} in a way that solves (7.11). G_{21} is entrusted with the job of seeing that $g_1 + g_2$ is maximized subject to the amount of the constraint $\underline{K}$ it is allocated and subject to the restrictions A_1, B_1. G_{22} is entrusted with a similar job with respect to $g_3 + g_4$. G_{11} has the problem of maximizing g_1 subject to the quantity of the constraint $\underline{K}$ allocated to it by its manager G_{21} and subject to the restrictions A_1. G_{1i} i=2,3,4 are similarly constructed.

We will now describe in a rough fashion the way this structure might operate, and we will assume that the reader can, from the treatment of 7.3, visualize the corresponding mathematical

development. We notice that the subsystems (G_{21}, G_{11}, G_{12}) and (G_{22}, G_{13}, G_{14}) are identical to the type of system treated in 7.2.

Step 1. G_3 allocates $\underline{K}$ to the second level units G_{21} and G_{22}.

Step 2. G_{21} using the amount of $\underline{K}$ allocated to them by G_3 pursue the course of action outlined in 7.2 until the amount of $\underline{K}$ initially allocated to G_{21} has been optimally apportioned to the first level systems.

Step 3. The optimal Lagrange multipliers from the first level systems are communicated to the second level units. The second level unit G_{21} then sends $\lambda_1^* + \lambda_2^*$ to the third level unit. G_{22} sends $\lambda_3^* + \lambda_4^*$ to the third level unit. The third level unit, in the same fashion as the second level unit in 7.2, using $\lambda_1^0 = \lambda_1^* + \lambda_2^*$ and $\lambda_2^0 = \lambda_3^* + \lambda_4^*$ changes his allocation of $\underline{K}$ between the two second level units.

This entire process repeats until the over-all optimum is reached.

Summary

In this chapter a "Decomposition Principle" for a class of non-linear programming problems was developed and numerical methods were proposed which converge to the optimum solution. The "Decomposition Principle" was interpreted as a control device for a class of decentralized organizational structures.

The extension of the basic model in several directions was mentioned and a description of its application to a multi-level structure ($m = 3$) was given.

The exact requirements for the application of the numerical techniques of this chapter are that if we wish to partition the optimization variables $m_1, \ldots m_n$ into two sets $m_1, \ldots m_k$ and $m_{k+1}, \ldots m_n$ then the objective functional must be of the form $g = g_1(m_1, \ldots m_k) + g_2(m_{k+1}, \ldots m_n)$ where g_i are concave, differentiable functions. The constraint functions $\bar{F}(\bar{m})$ must be of the form $\bar{A}(m_1, \ldots m_k) + \bar{B}(m_{k+1}, \ldots) \geq 0$ where $\bar{A}$ and $\bar{B}$ are vectors of concave, differentiable functions.

8. INDIRECT MODEL - INDIRECT INTERVENTION

AN EXAMPLE

Introduction

It was decided during the course of this research to devote
one period exclusively to an example which would model a hypothetical
real situation. It is the purpose of this chapter to illustrate
the application of some of the theoretical concepts introduced pre-
viously and to provide the reader with an actual numerical example
illustrating the utility of these ideas.

The situation we will consider is the following: consider
the operation of three separate water storage facilities (dams) all
located on the same river. These dams are assumed to be located in
an arid country such as Spain and the function of the system is to
store and allocate water for irrigation, industrial use, etc. The
hydroelectric aspects of dam operation are not considered here. The
control problem here is to allocate the water available to the system
to the system's customers over the dry period of the year (assumed
in this example to be six months) in such a way as to minimize some
measure of total cost. Each dam is assumed to have inputs from stream
flows. It is assumed that these inputs can be described as a
Markov process (since we are working in discrete time--a Markov
chain). This assumption was substantiated by J. D. C. Little[16]
who examined data from the inputs to Grand Coulee Dam and found
the stream flows to be autocorrelated and essentially Markovian.

Apparently seldom, if ever, in practice does the case of (time-wise) independent stream flows occur. Since we have 3 dams we will be dealing with 3 Markov chains $P_{ij}^{(1)}$, $P_{ij}^{(2)}$, $P_{ij}^{(3)}$. P_{ij} is the probability that if the stream flow was i units of water this period that j units will be forthcoming next period. The reader will notice that for the purpose of simplicity and due to the fact that we subsequently will need to employ a digital computer that we will assume the water is measured in discrete units rather than as a continuous variable.

For the purposes of visualization it may help the reader to compare this model with the Hungry Horse–Grand Coulee–Bonneville Dam system on the Columbia River. Although the Northwest is hardly an arid area and hydroelectric power is an important aspect of their operation, these dams are located in about the same relative positions as we visualize for our problem. We will denote our dams by number, i.e. #3 being farthest from the ocean, #2 next, etc. We assume that if water is released from one of the "higher" dams on the river this water will reach the "lower" dams in significantly less than one month. This is apparently justified for the Columbia system. The approximate transportation lag between Grand Coulee and Bonneville is 24 hours.

Each of the dams' customers (except possibly those of dam #1) are assumed to use all the water allocated to them, i.e. none is ever dumped back into the river.

Since we are concerned with an arid region it is assumed
that the capacity of each of the dams is adequate, i.e. dam capacity
is not a constraint.

The cost functions for each of the systems' customers is
assumed to be convex about some desired quantity. Graphically we
may illustrate this as

Fig. 8-1
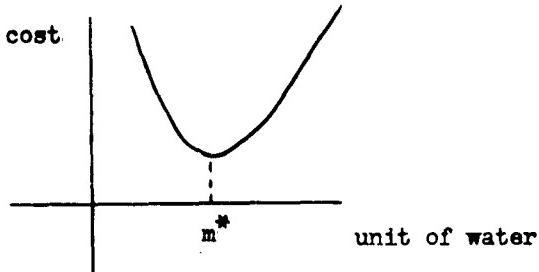
The cost functions are assumed to be independent from
period to period and fixed for each customer.

Each dam is assumed to have three customers, i.e. 9 in all.
We denote the cost functions for these customers by
$g_1 (m^1)$, $g_2 (m^2)$ ... $g_9 (m^9)$ where m^1 is the amount of water
allocated to customer i.

Let the amounts of water in the three dams in the current
period be z^1, z^2 and z^3 respectively. Let m^{10} be the quantity
of water released by dam 3 to dam 2, and m^{11} the amount released
from 2 to 1. We will indicate the scheduling period by subscripts,
e.g. m_1^9 is the amount of water allocated to customer #9 in the
first scheduling period. Let r_K^j be the amount of input (from
stream flows) to dam j in period K.

8.1   The Mathematical Problem

We can now formulate our control or optimization problem mathematically as follows:

$$\underset{m_i^1,\ldots m_i^{11}}{\text{Minimize}} \quad E\left\{\sum_{i=1}^{6}\sum_{j=1}^{9} g_j\,(m_i^j)\right\} \tag{8.1}$$

$$i = 1,\ldots 6$$

where E denotes the operation of mathematical expectation. The variables are subject to the following constraints.

$$0 \le m_1^1 + m_1^2 + m_1^3 + m_1^{10} \le z_1^1 \tag{8.2}$$

$$0 \le m_1^{11} + m_1^4 + m_1^5 + m_1^6 \le z_1^2 + m_1^{10}$$

$$0 \le m_1^7 + m_1^8 + m_1^9 \le z_1^3 + m_1^{11}$$

$$0 \le m_2^1 + m_2^2 + m_2^3 + m_2^{10} \le z_1^1 - m_1^1 - m_1^2 - m_1^3 + I_1^3 \tag{8.3}$$

etc.

Actually, of course, this is a very clumsy way to phrase this problem. A much neater forumulation is afforded by Dynamic Programming. In functional equation form the problem is:

$$f_N(z^1, z^2, z^3; I^1, I^2, I^3) = \min_{m^1, \ldots m^{11}} \left\{ \sum_{i=1}^{9} g_i(m^i) \right. \quad (8.4)$$

$$+ \sum_{\eta_1=0}^{N_1} \sum_{\eta_2=0}^{N_2} \sum_{\eta_3=0}^{N_3} P_{I^1, \eta_1}^{(1)} \quad P_{I^2, \eta_2}^{(2)} \quad P_{I^3, \eta_3}^{(3)}$$

$$f_{N-1}(z^1 - m^1 - m^2 - m^3 - m^{10} + \eta_1,$$

$$z^2 - m^4 - m^5 - m^6 - m^{11} + \eta_2 + m^{10},$$

$$\left. z^8 - m^7 - m^8 - m^9 + m^{11} + \eta_3; \eta_1, \eta_2, \eta_3) \right\}$$

$N = 1, \ldots C.$ $f_o(\ldots; \ldots) \equiv 0.$ The N_i are the orders of the respective Markov chains. The variables in (8.4) are subject to the following restrictions:

$$0 \leq m^1 + m^2 + m^3 + m^{10} \leq z^1 \quad (8.5)$$

$$0 \leq m^4 + m^5 + m^6 + m^{11} \leq z^2 + m^{10}$$

$$0 \leq m^7 + m^8 + m^9 \leq z^3 + m^{11}$$

The functional $f_N(\ldots; \ldots)$ can be interpreted as "the minimum expected cost for the system given that the initial state is $(z^1, z^2, z^3; I^1, I^2, I^3)$ and there are N decisions yet to make and further, that an optimal policy is followed."

It has often been said that "Dynamic Programming (as a conceptual device) can formulate many more problems than it can solve (numerically)." The reader has just seen such a case. First of all, the fact that the problem has six state variables leads to immense computer storage requirements, and secondly the maximization problem is imbedded in an 11 dimensional space which leads to staggering

run times for any conceivable search algorithm. As we will see subsequently this problem for 6 periods using a very coarse grid approximation would entail storage requirements of on the order of 10^7 locations and run times on the Burroughs 220 of on the order of 10^6 years. Clearly this is not a problem where we can wait for a bigger, faster machine. If we really want a numerical solution to this problem we will have to be a little more ingenious.

The general approach we will use here will be the same as is employed in the preceeding chapters. We will first decouple the system and solve three separate control problems and then recouple the system into a 2 level- 4 goal system where the second level controller is responsible for recoupling in such a manner that the interaction between the first level systems is compensated.

8.2 Synthesis of the 2 ℓ- 4 g System

The formulation of the problems for the three first level systems is a simple procedure. The manager of each dam is responsible for the solution of a Dynamic Programming problem for his dam, neglecting the possibility that he may receive water from the dam above or be required to spill water to the dam below. For example, dam #3 solves the following problem.

$$f_N^{(3)}(z^3, I^3) = \min_{m^1,m^2,m^3} \left\{ \sum_{i=1}^{3} g_i(m^1) + \sum_{\eta^3=0}^{N^3} P_{I^3,\eta^3}^{(3)} \cdot \right.$$

$$\left. f_{N-1}^{(3)}\left(z^3 - \sum_{i=1}^{3} m^1 + \eta^3, \eta^3\right) \right\} \text{ where } 0 \leq \sum_{I} m^1 \leq z^3$$

$$f_0^{(3)}(\cdot, \cdot) \equiv 0 \qquad\qquad N = 1,\ldots 6.$$

A similar expression holds for the optimization problems to be solved by dams 1 and 2. The $f_K^{(j)}(z^j, I^j)$ are to be tabulated for all three dams ($j = 1, 2, 3$) and all six periods ($K = 1,\ldots 6$) and all feasible values of z^j and I^j within the expected range of operation. These solutions are then transmitted to the second level unit. Only the f values are required and not the m^{i*}.

Now we will consider the functions of the second level unit. Consider $f_N^1(z^1, I^1) - f_N^1(z^1 + 1, I^1)$. This is, by the definition of f_N^1, the expected marginal value of water to dam 1 given that N decisions have yet to be made, z^1 units of water are now in the dam, and the last period's input was I^1 units. By similarly examining $f_N^{*1}(z^1 + 2, I^1)$, $f_N^1(z^1 + 3, I^1)$, etc., we could by polynomial fit or similar method construct a function which is dam 1's marginal utility function for extra water. Assume this function has been constructed. We will denote it as $G_1(m^{11})$ since m^{11} is where extra water, if any, will be forthcoming.

Now dam 2 is asked to solve the following problem:

find

$$f_N^{2^t} (z^2, I^2) = \min_{m^4, m^5, m^6, m^{11}} \left\{ \sum_{4=1}^{6} g_i (m^i) - G_1 (m^{11}) \right. \quad (8.7)$$

$$\left. + \sum_{\eta_2=0}^{N_2} P_I^{(2)} , \eta_2 \quad f_{N-1}^2 (z^2 - \sum_{4=1}^{6} m^i - m^{11} + \eta_2, \eta_2) \right\}$$

Note that in the right side of (8.7) we are using f_{N-1}^2 and not

$f_{N-1}^{2^t}$, hence this is a single optimization problem and not N prob-

lems, because f_{N-1}^2 has already been tabulated.

Now exactly the same process is repeated for dam 2 that

was carried for dam 1, i.e. $f_N^t (z^2, I^2)$, $f_N^t (z^2 + 1, I^2) \ldots$ are

used to construct $G_2 (m^{10})$. Dam 3 is now asked to solve the

problem: find

$$f_N^3 (z^3, I^3) = \min_{m^1, m^2, m^3, m^{10}} \left\{ \sum_{i=1}^{3} g_i (m^i) - G_2 (m^{10}) \right. \quad (8.8)$$

$$\left. + \sum_{\eta_3=0}^{N_3} P_I^{(3)}{}_3, \eta_3 \quad f_{N-1}^3 (z^3 - \sum_{i=1}^{3} m^i - m^{10} + \eta_3, \eta_3) \right\}$$

The results of this optimization are then implemented, i.e. dam 3

gives m^{1*} units of water to customer 1, m^{2*} units to customer 2,

etc., and m^{10*} units are spilled from dam 3 to dam 2.

Dam 2 now finds the optimal values of m^4, m^5, m^6 and m^{11}

corresponding to $f_N^{2^t} (z^2 + m^{10}, I^2)$ and implements them.

Dam 1 then finds the optimal values of m^7, m^8, m^9 corresponding to f_N^1 ($z^1 + m^{11}$, r^1). These are then implemented. The inputs from the streams are assumed to occur at this point, and the process repeats.

The function of the second level unit here is the administration of the recoupling process and specifically the construction of the functions G_1 and G_2 from the separate Dynamic Programming solutions.

Although the recoupling process is lengthy to describe it is quite efficient computationally, as we will see.

8.3 Computational Results

The entire process described above was programmed for the Burroughs 220 Computer. The following simplifications were introduced to cut "run" times to a minimum.

1. g_1 (m^1) = $(m^1 - 1)^2$, g_2 (m^2) = $(m^2 - 2)^2$,

   g_3 (m^3) = $(m^3 - 3)^2$

2. g_4 and g_7 have the same form as g_1, g_5 and g_8 have the same form as g_2, and g_6 and g_9 have the same form as g_3.

3. All three dams were assumed to have identical input Markov chains.

$$
P_{ij} = \begin{bmatrix} .3 & .3 & .3 & .1 \\ .2 & .4 & .2 & .2 \\ .2 & .2 & .4 & .2 \\ .2 & .3 & .4 & .1 \end{bmatrix}
$$

4. A simple 2 point quadratic fit is used to obtain the G_i using f_N^t (z, I) and f_N^t $(z + 1, I)$ and f_N^t $(z + 2, I)$.

The input process for each of the three dams was simulated for each of six periods and these inputs were used to simulate the decision-making process. The Dynamic Programming solution took roughly 10 minutes on the 220 and thereafter the complete simulation of the six decision periods, including the random input generation, averaged about two minutes. A total of 11 complete six-period simulations were run, i.e. 66 decision periods.

The following modifications could have been accommodated and still keep the D. P. calculations and one simulation to on the order of one hour run time.

1. 3 separate D. P. programs instead of one.
2. Larger Markov chains, up to possibly 8 x 8.
3. Non-stationary cost functions and/or Non-stationary Markov chains.
4. More sophisticated fitting procedures to obtain the G_i.
5. More periods than six.

For example, the P_{11} element tells us that if we received 0 units of water last month we will receive 0 units this month with probability .3 .

The method of search employed in the program was simple enumerative
search point by point. A more sophisticated technique such as
"Steepest Descent" would make the computation of much finer grid
sizes possible.

We will now compare the storage requirements and "run" times
of this simulation to an estimate of the requirements for the solu-
tion of (8.4). We will assume that the average grid size per state
variable was 10, i.e. the solution was calculated for 10 values of
each of the state variables. We assume that the number of scheduling
periods is six. A little arithmetic will convince the reader that
the storage requirements for the solution of (8.4) is on the order
of 10^7 memory locations, while for our routine the requirements
can be kept to well under 10^4. Therefore if the cost of computation
goes up linearly with the number of required storage locations the
ratio of the computation cost of (8.4) versus the procedure outlined
in 8.2 is on the order of 10^3, i.e. the solution of (8.4) would be
at least a thousand times were costly than the approach outlined
in this chapter. Actually, there are no existing computing facilities
with the size required by (8.4), and hence its solution is unfeasible,
regardless of cost.

We now turn to a consideration of the performance of our
method. The following procedure was used to find an upper bound
for the possible performance. If one knew at the beginning of period
1 exactly how much water would collectively be available to the dams,

and if all that water were available in dam 3 at the beginning of
period 1, then it could be allocated over the six periods in such a
way that no better allocation would be possible. This condition is,
of course, unachievable, but it does yield a lower bound on achievable
costs. This cost is called the "Perfect" cost and is compared with
the costs obtained by our algorithm. A complete tabulation of
results of the 11 simulations follows.

Run # 1

Period	Amount of Input:			
	Dam 1	Dam 2	Dam 3	Cost
1	1	1	1	27
2	0	1	2	27
3	0	1	2	27
4	0	2	2	27
5	2	2	0	29
6	3	2	2	11
Total	6	9	9	148
"Perfect"				144
% dev. from "Perfect"				2.9%

Run # 2

Period	Amount of Input:			
	Dam 1	Dam 2	Dam 3	Cost
1	1	1	1	27
2	2	1	2	27
3	2	2	3	27
4	3	2	3	9
5	0	3	2	21
6	1	0	2	14
Total	9	9	13	125
"Perfect"				123
% dev. from "Perfect"				1.60%

Run #3

Period	Amount of Input:			
	Dam 1	Dam 2	Dam 3	Cost
1	1	1	1	27
2	2	1	2	27
3	3	0	1	24
4	2	2	2	21
5	2	3	1	18
6	2	2	3	8
Total	12	9	10	125
"Perfect"				123
% dev. from "Perfect"				1.6%

Run #4

Period	Amount of Input:			
	Dam 1	Dam 2	Dam 3	Cost
1	1	1	1	27
2	0	1	2	27
3	2	3	0	32
4	2	2	0	23
5	2	0	1	24
6	2	0	0	31
Total	9	7	4	164
"Perfect"				156
% dev. from "Perfect"				5.1%

Run # 5

Period	Amount of Input:			
	Dam 1	Dam 2	Dam 3	Cost
1	1	1	1	27
2	1	2	2	27
3	2	1	2	27
4	2	1	3	15
5	2	1	1	24
6	2	1	0	18
Total	10	7	9	138
"Perfect"				138
% dev. from "Perfect"				0.0%

Run # 6

Period	Amount of Input:			
	Dam 1	Dam 2	Dam 3	Cost
1	0	0	2	37
2	0	1	2	27
3	0	1	2	27
4	0	2	2	27
5	2	2	0	24
6	3	2	2	12
Total	5	8	10/23	154
"Perfect"				147
%dev. from "Perfect"				4.55%

Run # 7

Period	Amount of Input:			
	Dam 1	Dam 2	Dam 3	Cost
1	0	0	2	37
2	2	0	2	27
3	2	2	0	29
4	1	0	3	24
5	0	1	3	24
6	0	1	1	21
Total	5	4	11/20	162
"Perfect"				156
%dev. from "Perfect"				3.70%

Run # 8

Period	Amount of Input:			
	Dam 1	Dam 2	Dam 3	Cost
1	0	0	2	37
2	1	2	2	27
3	2	1	2	27
4	2	1	3	12
5	2	1	1	24
6	2	1	0	18
Total	9	6	10/25	145
"Perfect"				141
%dev. from "Perfect"				2.78%

Run # 9

Period	Amount of Input:			
	Dam 1	Dam 2	Dam 3	Cost
1	0	0	2	37
2	2	1	2	27
3	2	2	3	24
4	3	2	3	9
5	0	3	2	21
6	1	0	2	14
Total	8	8	14/30	132
"Perfect"				126
%dev. from "Perfect"				4.55%

Run # 10

Period	Amount of Input:			
	Dam 1	Dam 2	Dam 3	Cost
1	0	0	2	37
2	2	I	2	27
3	3	0	1	21
4	2	2	2	21
5	2	3	1	18
6	2	2	3	8
Total	11	8	11/30	132
"Perfect"				126
%dev. from "Perfect"				4.55%

Run # 11

Period	Amount of Input:			
	Dam 1	Dam 2	Dam 3	Cost
1	0	0	2	37
2	0	0	2	27
3	2	2	0	37
4	2	2	0	26
5	2	0	1	24
6	2	0	0	31
Total	8	4	5/17	182
"Perfect"				167
%dev. from "Perfect"				8.25%

For the first five runs the average percent deviation from
"Perfect" is 2.2%. For the last six runs the average percent deviation from "Perfect" is 4.73%. The performance of the solution
to (8.4) must lie somewhere between the performance of our algorithm
and "Perfect," and hence our algorithm performs very well indeed.
Our algorithm's performance is less than 2.2% from the optimum
achievable performance for the first five runs and less than 4.73%
from the optimum in the last six runs.

Summary

In this chapter we have presented a complete numerical
analysis of an example of a $2\,\ell-4\,g$ system where the second level
unit employs an indirect model of the first level systems below it
and intervenes in the affairs of the first level systems via
indirect intervention. The model employed by the second level is
indirect because it requires only a knowledge of $f_j^{(i)t}(z, I)$ instead

of a complete knowledge of the dynamics of system i. The inter-
vention is indirect because the operation of the second level unit
consists of inserting the factors G_1 and G_2 in the performance
functionals of dam 2 and dam 3 respectively.

The optimization problem was solved and the total opera-
tion was simulated over 66 decision periods. A lower bound on the
costs is generated and it is shown that the algorithm's simulated
performance usually falls within 3 to 5% of this lower bound.

It is also shown that the (computation) cost advantage of this algorithm over a complete Dynamic Programming solution is on the order of a thousand to one.

9. CONCLUSIONS

In this thesis we have investigated a number of "organizational" systems or $m\ell - ng'$ systems where in each case the over-all system task is the solution of a large multi-variable optimization problem. Our principle effort was directed toward the development of efficient "coordination" and "modeling" techniques for the higher level or "managerial" units.

In Chapter 3 we introduced the concept of direct intervention, and we found that for certain classes of problems the performance of the synthesized $2\ell - 3g'$ system was the best achievable and that for all the deterministic models of Chapter 3 the direct intervention approach was better than simply neglecting the interaction between the first level units. The limitations of this method in the static case are that it requires special physical conditions for its implementation, it places heavy information processing loads on the managerial unit, and it does not readily extend to more than two levels.

In Chapter 4 we introduced indirect intervention. Again this approach yields better performance than one obtains by reticulating the system and then neglecting the interaction between reticulated parts. This method requires no special conditions for its implementation, and further it significantly reduces the computational load on the managerial unit. However, this method does reduce the controllability of the over-all system as viewed from the point

of view of the highest level unit. This in itself is no handicap because it is possible for the second level unit to have very few controllable variables and still be able to steer the performance of the over-all system to the optimum.

In Chapter 5 we found that the direct intervention approach applied to a broad class of linear dynamic systems leads us to the conclusion that in this case the $2\ell - 3g$ system is equivalent to a $1\ell - 2g$ system.

In Chapter 6 we employed indirect intervention in the dynamic case and showed that it can be used to significantly reduce the dimensionality of the search problem. We also were able to demonstrate that prediction can arise in an "organization" purely as a function of the structure of the system. The planning function in a multi-level system need not be linked to environmental uncertainty at all, but may be built into the system as a control device.

In Chapter 7 we introduced the concept of the indirect model. In Chapters 3, 4, 5 and 6 the second level unit is assumed to have complete knowledge of the entire system and its environment, i.e. it has a direct model of the system below it. In Chapter 7 only one number is required by the second level system from each lower level unit in any one decision period in order that the second level unit be able to steer the complete system to the optimal over-all performance. The feasibility of extending this type of model

to more than two levels was investigated, and one class of three
level, seven goal systems was discussed in some detail.

In Chapter 8 we applied the concepts of indirect interven-
tion and indirect modeling to synthesize a $2\ell - 4g$ system for the
control of a 3-dam water storage system. Numerical results were
obtained and the quality of performance as well as computational
costs were discussed.

BIBLIOGRAPHY

1. Adorno, D. "Optimum Control of Certain Linear Systems with Quadratic Loss", Information and Control 5, 1-12 (1962).

2. Arrow, K., Hurwiez, L., Usawa, H. Studies in Linear and Non-Linear Programming, Stanford, Calif., Stanford University Press, 1958.

3. Bellman, R. Adaptive Control Processes - A Guided Tour, Princeton, N. J., Princeton University Press, 1961.

4. Bellman, R. Dynamic Programming, Princeton, N. J., Princeton University Press, 1957.

5. Bellman, R. Introduction to Matrix Analysis, New York, McGraw-Hill, 1960.

6. Brockett, R. and Mesarovic, M. D. Output Controllability of Multivariable Systems (1962). (to appear)

7. Chang, S. Synthesis of Optimum Control Systems, New York, McGraw-Hill, 1961.

8. Dantzig, G. and Wolfe, P. "Decomposition Principle for Linear Programs", Op. Res. 8, 101 (1960).

9. Doob, J. Stochastic Processes, New York, John Wiley, 1953.

10. Halmos, P. Finite Dimensional Vector Spaces, Princeton, N. J., Princeton University Press, 1942.

11. Jury, E. Sampled Data Control Systems, New York, John Wiley, 1958.

12. Kalman, D. R. E. "On the General Theory of Control Systems", Proceedings of the First International Congress of the International Federation of Automatic Control, London, Butterworth, 1961.

13. Karlin, S. Mathematical Methods and Theory in Games, Programming and Economics, Reading, Mass., Addison-Wesley, 1959.

14. Kirchmayer, A. Economic Control of Interconnected Systems, New York, John Wiley, 1959.

15. Kirchmayer, A. Economic Operation of Power Systems, New York, John Wiley, 1958.

16. Little, J. "The Use of Storage Water in a Hydroelectric System", J. Operations Research Soc. Am. 3, 187 (1955).

17. Mesarovic, M. D. and Eckman, D. P. On Some Basic Concepts of the General Systems Theory, Namur, Belgium, August 1961.

18. Mesarovic, M. D. A General Systems Approach to Organizational Theory, Brussels, Belgium, August 1961.

19. Middleton, D. An Introduction to Statistical Communication Theory, New York, McGraw-Hill, 1960.

20. Orner, P. "On the Measure of Interaction", Systems Research Center Progress Report No. 1-7, Case Institute of Technology (1961)

21. Pruzan, P. Ph.D. Thesis, Case Institute of Technology (1963). (to appear)

22. Scheffe, A. The Analysis of Variance, New York, John Wiley, 1959.

23. Taylor A. Introduction to Functional Analysis, New York, John Wiley, 1958.

24. Whinston, A. "Price Guides in Decentralized Organization", Conference on Organization Theory, Carnegie Institute of Technology (1962).

APPENDIX A

In this Appendix we will prove several results relating to matrix analysis and the theory of operators. It is assumed that the reader is familiar with matrix algebra and real analysis.

We will define the following notation

$\bar{m}$ is an n-component constant vector.

$\bar{\epsilon}$ is an n-component random vector with

expected value $E\bar{\epsilon} = \bar{\mu}$.

A is an n x n constant vector.

Theorem A-1. $E(\bar{m}^T A \bar{\epsilon}) = \bar{m}^T A \bar{\mu}$

Proof: Let $\bar{\bar{m}}^T = \bar{m}^T A = \left[\bar{\bar{m}}_1, \bar{\bar{m}}_2 \ldots \bar{\bar{m}}_n \right]$

$\bar{m}^T A \bar{\epsilon} = \bar{\bar{m}}_1 \epsilon_1 + \bar{\bar{m}}_2 \epsilon_2 + \ldots \bar{\bar{m}}_n \epsilon_n$

$E(\bar{m}^T A \bar{\epsilon}) = \bar{\bar{m}}_1 \mu_1 + \bar{\bar{m}}_2 \mu_2 + \ldots \bar{\bar{m}}_n \mu_n = \bar{\bar{m}}^T \cdot \bar{\mu}$ Q.E.D.

Theorem A-2. $\bar{m}^T A \bar{\mu} = \bar{\mu}^T A^T \bar{m}$

As above, let $\bar{\bar{m}}^T = \bar{m}^T A$ $\bar{\bar{m}} = A^T \bar{m}$

Therefore we can write the theorem as $\bar{\bar{m}}^T \bar{\mu} = \bar{\mu}^T \bar{\bar{m}}$.

$\bar{\bar{m}}^T \bar{\mu} = \left[\bar{\bar{m}}_1, \bar{\bar{m}}_2, \bar{\bar{m}}_3 \ldots \bar{\bar{m}}_n \right] \cdot \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$

$= \bar{\bar{m}}_1 \mu_1 + \bar{\bar{m}}_2 \mu_2 \ldots \bar{\bar{m}}_n \mu_n$

but $\bar{\mu}^T \bar{\bar{m}} = \begin{bmatrix} \mu_1 & \cdots & \mu_n \end{bmatrix} \begin{bmatrix} \bar{\bar{m}}_1 \\ \vdots \\ \bar{\bar{m}}_n \end{bmatrix}$

$= \mu_1 \bar{\bar{m}}_1 + \mu_2 \bar{\bar{m}}_2 + \cdots \mu_n \bar{\bar{m}}_n = \bar{\bar{m}}^T \bar{\mu}$ \qquad Q.E.D.

Theorem A-3. $\dfrac{d}{d\bar{m}} (\bar{m}^T A \bar{m}) = 2 A \bar{m}$ \qquad if $A^T = A$

Proof: By $\dfrac{d}{d\bar{m}}$ we mean the vector of derivatives $\begin{bmatrix} \dfrac{d}{d m_1} \\ \vdots \\ \dfrac{d}{d m_n} \end{bmatrix} \cdot \bar{m}^T A \bar{m} =$

$m_1 a_{11} m_1 + m_1 a_{12} m_2 + \cdots m_n a_{nn} m_n \cdot \dfrac{d}{dm_1} (\bar{m}^T A \bar{m}) = 2 a_{11} m_1 +$

$a_{12} m_2 + a_{21} m_2 + a_{13} m_3 + a_{31} m_3 + \cdots a_{1n} m_n + a_{n1} m_n$ \qquad but

$a_{ij} = a_{ji}$ \qquad hence $\dfrac{d}{d m_1} (\bar{m}^T A \bar{m}) = 2 a_{11} m_1 + 2 a_{12} m_2 + \cdots 2 a_{1n} m_n$

$= 2 A_1 \cdot \bar{m}$ \qquad where $A_1 = \begin{bmatrix} a_{11}, & a_{12}, & \cdots & a_{1n} \end{bmatrix}^T$

Similarly $\dfrac{d}{d m_i} (\bar{m}^T A \bar{m}) = 2 A_i \cdot \bar{m}$ \qquad $\dfrac{d (\bar{m}^T A \bar{m})}{d \bar{m}} = 2 \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix} \bar{m}$

$= 2 A \bar{m}.$

Example. minimize $E (\bar{y}^T B \bar{y})$ where $\bar{y} = A \bar{m} + \bar{\epsilon}$ and $B = B^T$

$$\bar{y}^T B \bar{y} = (A \bar{m} + \bar{\epsilon})^T B (A \bar{m} + \bar{\epsilon}) = \bar{m}^T A^T B A \bar{m}$$

$$+ \bar{m}^T A^T B \bar{\epsilon} + \bar{\epsilon}^T B A \bar{m} + \bar{\epsilon}^T B \qquad \text{by using Theorem A-2 we}$$

reduce $\bar{y}^T B \bar{y} = \bar{m}^T A^T B A \bar{m} + 2 m^T A^T B \bar{\epsilon} + \bar{\epsilon}^T$ therefore by A-1,

$$E (\bar{y}^T B \bar{y}) = \bar{m}^T A^T B A \bar{m} + 2 m^T A^T B \bar{\mu} + E(\bar{\epsilon}^T B \bar{\epsilon})$$

$$\frac{d}{d \bar{m}} E (\bar{y}^T B \bar{y}) = 2 A^T B A \bar{m} + 2 A^T B \bar{\mu} \quad \text{since } E (\bar{\epsilon}^T B \bar{\epsilon}) \text{ is a}$$

constant. If we equate the vector derivative to 0 we have

$$2 A^T B (A \bar{m} + \bar{\mu}) = 0 \qquad \text{If } A \neq 0 \text{ and } B \neq 0 \qquad \bar{m}^* = - A^{-1} \bar{\mu}$$

where m^* is the optimizing value of the vector $\bar{m}$.

Now we will take the relevant portion of a theorem from Halmos[10] which establishes the fact that we can represent any linear operator mapping a finite dimensional linear vector space $\underline{X}$ into itself as a finite matrix.

Theorem A-4. The correspondence (established by means of an arbitrary coordinate system $\chi = (x_1, \ldots x_n)$ of the n-dimensional vector space $\underline{X}$) between all linear transformations A on $\underline{X}$ and all matrices α_{ij} described by $A x_j = \sum_i \alpha_{ij} x_i$ is an isomorphism.

Proof: The proof can be found in Halmos[ibid.] We state without proof that it also preserves the operation of multiplication (operator-matrix) and inversion.

In this section we give the definitions and preliminaries which are the necessary background for Theorem 4.3 of Chapter 4. The definitions and theorems stated here are taken from selected portions of Taylor.[21]

Definitions.

1. The space of all continuous linear operators mapping the linear space $[X]$ into itself will be denoted by X .

2. If X and Y are normed linear spaces, we define the "norm" of the operator A mapping X into Y as
$$||A|| = \sup_{||x|| \leq 1} ||A\,x|| \quad x \in X \qquad \text{where}$$
$||x||$ is the norm of the vector x, $||A\,x||$ is the norm of the vector $A\,x \in Y$.

3. If λ is such that the range of $\lambda I - T$ is dense in X where T has both its domain and range in X and I is the identity operator, and if $\lambda I - T$ has a continuous inverse, we say λ is in the resolvant set of T ; this set of values of λ is denoted by $\rho(T)$. All scalar values of λ not in $\rho(T)$ comprise the set called the spectrum of T ; it is denoted by $\sigma(T)$.

4. A space X is a Banach space if it is a complete, normed, linear space.

The Theorem 4.3 of Chapter 4 is proved in Taylor.[ibid.] It is, however, somewhat more powerful than we require, and hence we will state and prove that portion of the theorem that we require.

Let $\left[X\right]$ be the space of all bounded linear operators mapping the Banach space X into itself.

Theorem A-5. If $T \in \left[X\right]$ and $|\lambda| > ||T||$ and if $(\lambda I - T)^{-1}$ exists and is continuous, then

$$(\lambda I - T)^{-1} y = \sum_{1}^{\infty} \lambda^{-n} T^{n-1} y \quad \text{for all } y \text{ in the range}$$

of $\lambda I - T$.

Proof: If $y = (\lambda I - T) x$ then

$$x = \lambda^{-1} y + \cdots \quad \lambda^{-n} T^{n-1} y + \lambda^{-n} T^{n} x$$

we have the results because $\lambda^{-n} T^{n} x \longrightarrow 0$ when $|\lambda| > ||T||$ hence the series converges geometrically in the norm of the space X.

APPENDIX B

In this Appendix we present the proofs of the Dynamic Programming theorems used in Chapters 5 and 6.

Consider the functional equation

$$S_{k+1}(\bar{c}) = \max_{\bar{m}_{N-k}} E\left\{ (A\,\bar{c} + \bar{m}_{N-k} + \bar{\epsilon}_{N-k})^T B (A\,\bar{c} + \bar{m}_{N-k} + \bar{\epsilon}_{N-k}) \right. \tag{B.1}$$

$$\left. - \bar{m}_{N-k}^T C\,\bar{m}_{N-k} + S_k (A\,\bar{c} + \bar{m}_{N-k} + \bar{\epsilon}_{N-k}) \right\}$$

where $E(\bar{\epsilon}_{N-k}) = \bar{\mu}_{N-k}$ and $S_0(\bar{c}) = 0$ for all $\bar{c}$.

$$S_1(\bar{c}) = \max_{\bar{m}_N} E\left\{ (A\,\bar{c} + \bar{m}_N + \bar{\epsilon}_N)^T B (A\,\bar{c} + \bar{m}_N + \bar{\epsilon}_N) \right. \tag{B.2}$$

$$\left. - \bar{m}_N^T C\,\bar{m}_N \right\}$$

From (B.2) by taking the expectation and derivative we find

$$\bar{m}_N^* = H_N A\,\bar{c} + G_N \quad \text{where} \quad H_N = (C - B)^{-1} B$$

and $G_N = (C - B)^{-1} B\bar{\mu}_N$.

We therefore state and prove the following theorem.

Theorem B-1. The structure of the optimal policy is

$m_{N-k}^* = H_{N-k} A\,\bar{c} + G_{N-k}$ and the solution to (B.1) is

$S_k(\bar{c}) = c^T A^T \psi_{N-k+1} A\,\bar{c} + 2\,\delta_{N-k+1} A\,\bar{c} + \phi_{N-k+1}$ where:

$$H_{N-k} = (C - B_{N-k})^{-1} B_{N-k}, \qquad B_{N-k} = B + A^T \psi_{N-k+1} A$$

$$H_N = (C - B)^{-1} B \qquad G_N = (C - B)^{-1} B \bar{\mu}_N$$

$$\psi_{N-k} = (I + H_{N-k})^T B_{N-k} (I + H_{N-k}) - H_{N-k}^T C H_{N-k}$$

$$\delta_{N-k} = (I + H_{N-k})^T B_{N-k} (\bar{\mu}_{N-k} + G_{N-k}) - H_{N-k}^T C G_{N-k} + A^T \delta_{N-k+1}$$

$$\delta_{N+1} = 0, \qquad \psi_{N+1} = 0, \qquad \phi_{N+1} = 0$$

$$\phi_{N-k} = \phi_{N-k+1} + 2 \delta_{N-k}^T A (G_{N-k} + \bar{\mu}_{N-k}) +$$

$$E \left\{ (\epsilon_{N-k} + G_{N-k})^T B_{N-k} \cdot (\epsilon_{N-k} + G_{N-k}) \right\}$$

Proof: We proceed by induction. We have previously demonstrated the hypothesis to be true for $k = 0$ by showing $\vec{m}_N = H_N A \bar{c} + G_N$ and the hypothesis on $S_0 (\bar{c})$ is satisfied as can be verified by inspection.

Let the induction variable be ℓ . If we assume that hypothesis is true for all $\ell \leq k-1$ we have

$$S_k (\bar{c}) \; \bar{c}^T A^T \psi_{N-k+1} A \bar{c} + 2 \delta_{N-k+1} A \bar{c} + \phi_{N-k+1}$$

$$S_{k+1} (\bar{c}) = \max_{\bar{m}_{N-k}} E \left\{ (A \bar{c} + \bar{m}_{N-k} + \epsilon_{N-k})^T B (A \bar{c} + \bar{m}_{N-k} + \epsilon_{N-k}) \right.$$

$$- \bar{m}_{N-k}^T C \bar{m}_{N-k} + (A \bar{c} + \bar{m}_{N-k} + \epsilon_{N-k})^T A^T \psi_{N-k+1} A (A \bar{c} + \bar{m}_{N-k} + \epsilon_{N-k})$$

$$\left. + 2 \delta_{N-k+1}^T A (A \bar{c} + \bar{m}_{N-k} + \epsilon_{N-k}) + \phi_{N-k+1} \right\}$$

Taking the expectation and derivative we find

$$m^*_{N-k} = (C - B_{N-k})^{-1} \; B_{N-k} \; (A \; \bar{c} + \bar{\mu}_{N-k}) + A^T \; \mathcal{S}_{N-k+1}$$

and now an examination of H_{N-k} and G_{N-k} verifies the hypothesis on m^*_{N-k}. Substitution of this result into $S_{k+1} \; (\bar{c})$ verifies the structure of $S_k \; (\bar{c})$.

Theorem B-2. The solution of the functional equation

$$S_{k+1} \; (\bar{c})' = \min_{\bar{m}_{N-k}} \; E \left\{ (A \; \bar{c} + \bar{m}_{N-k} + \bar{\epsilon}_{N-k}) \; B \; (A \; \bar{c} + \bar{m}_{N-k} + \bar{\epsilon}_{N-k}) \right.$$

$$\left. + \; \bar{d}^T_{N-k} \; \bar{m}_{N-k} + S_k \; (A \; \bar{c} + \bar{m}_{N-k} + \bar{\epsilon}_{N-k}) \right\}$$

where $\bar{d}_i$ is a sequence of constant vectors is

$$S_{N-k} \; (\bar{c}) = \mathcal{S}_{N-k} - 2 \; \bar{d}^T_{N-k+1} \; A \; \bar{c} \qquad \text{and the optimal policy is}$$

$$\bar{m}^*_{N-k} = B^{-1} \; (A^Z \; \bar{d}_{N-k+1} - \bar{d}_{N-k}) - (A \; \bar{c} + \bar{\mu}_{N-k})$$

where $\mathcal{S}_k = \mathcal{S}_{k-1} + E \; (\alpha_{N-k+1} + \bar{\epsilon}_{N-k+1} - \bar{\mu}_{N-k+1})^T \; B \; (\alpha_{N-k+1}$

$$+ \; \bar{\epsilon}_{N-k+1} - \bar{\mu}_{N-k+1}) \; + \; 2 \; \bar{d}^T_{N-k+1} \; (\alpha_{N-k+1} - \bar{\mu}_{N-k+1})$$

$$- \; 2 \; \bar{d}_{N-k+2} \; A \; \alpha_{N-k+1}$$

and $\alpha_{N-k} = B^{-1} \; (A^T \; \bar{d}_{N-k+1} - \bar{d}_{N-k})$

Proof: The proof proceeds by induction in exactly the same form as Theorem B-1. and hence we will omit it.

APPENDIX C

This Appendix formulates a possible area of application for the models of Chapters 3, 4, 5 and 6. It is our intent to indicate the general areas of application and to interpret the features of the models of Chapters 3 through 6 in the language of the application area; however, we will not concern ourselves with the specific details of applying the results.

In this Appendix we will interpret the vectors $\bar{y}$ and $\bar{y}_t$ as output vectors of a system which express the deviation in the system outputs from a fixed operating point, i.e. $\bar{y} = \bar{z} - \bar{z}_o$ where $\bar{z}_o$ is the vector of fixed operating points and $\bar{z}$ is the actual system output. The $\bar{\varsigma}$ and $\bar{\varsigma}_t$ are to be interpreted as vectors of disturbances which appear in the outputs and arise from causes external to the system. The $\bar{m}$ and $\bar{m}_t$, as before, are vectors of manipulated variables.

An example of a system where the disturbances appear directly across the outputs of the system as they do in the models of these chapters is in any electrical generating system. A class of practical situations where these models might find application are in hydro-electric generating stations where there are several turbines operating off a single penstock and where increases in load on one generator affect the speed of the other generators.

The System Equations

In Chapters 3 and 4 we deal with system models of the type $\bar{y} = A\,\bar{m} + \bar{\epsilon}$. Here we assume that $\bar{y}$ is a function of time but that it depends solely on $\bar{m}$ and $\bar{\epsilon}$ and not on its own past as do the $\bar{y}_t$ of Chapters 5 and 6. We interpret $\bar{m}$ in this context as being the magnitude of the penstock value openings to the respective turbines. $\bar{m}$ is assumed to be a function of time and it is assumed that the variations of $\bar{m}$ are sufficiently small that the inherent non-linearities of the system do not influence the model.

A is the "system matrix" which symbolically transforms penstock value openings into generator outputs. A is assumed to be constant. We will assume that the disturbances $\bar{\epsilon}$ are functions of time but known, i.e. the fluctuations in load have been previously observed and are periodic. We make this assumption here to avoid having to deal with random functions.

The dynamic models of Chapters 5 and 6 can arise in two different contexts. First we might assume that the system had significant internal dynamics, i.e. that the internal inertia of the system, for example, makes the outputs at time $t+1$ dependent on the outputs at time t. Alternatively we might assume that we are dealing with two different dams located on the same river so that the water available to the lower dam is dependent on the discharge of the higher dam.

The Performance Functionals

Two of the most important goals of a power generating facility[15] are to maintain constant output levels independent of the load and to minimize the transmission losses which are a function of the output levels. We will assume that the generating facility appraises the cost of deviating from the fixed operating point $\bar{z}_0$ as being proportional to the square of the deviation. We can state the first goal mathematically as minimize $\bar{y}^T B \bar{y}$ where B is a positive definite matrix. It has been shown in the literature[15] that the transmission losses go up also proportional to the square of the outputs. Hence we can define the control problem for the system as

$$\underset{\overline{m}}{\text{minimize}} \quad \bar{y}^T B \bar{y} + (\bar{y} + \bar{z}_0)^T C (\bar{y} + \bar{z}_0) \; .$$

(C is positive definite)

Direct Intervention

There are at least two ways to interpret the terms $A_{ij} \lambda^j$ which occur in the output equations of Chapters 3 and 5. One way is to assume that λ^j is a control signal and that the direct intervention procedure requires the capability of tapping into the generator speed controls at an appropriate point in such a way as to introduce λ^{-1} into A_{21} and λ^{-2} into A_{12}. A second way is to assume that the total term, e.g. $A_{12} \lambda^{-2}$, is inserted directly into the output. In this method then one takes the vector signal λ^{-2}, amplifies it by the matrix A_{12} and inserts the resultant directly into the outputs.

Current hydroelectric facilities generally possess the ability to implement this procedure directly due to the presence of lines which allow them to tap power from a neighboring power system. These lines are known as "tie lines".[14]

Indirect Intervention

This mode of intervention is concerned solely with the computational aspects of control and hence has no counterpart in the physical system.

APPENDIX D

In this Appendix we will define "controllability" in a manner which is suitable for our requirements and indicate the other uses of the term which are prevalent in the literature.

The concept of controllability, broadly stated, is based on the following:[12] Given a system, a set of required outputs, and a set of available manipulated variables, is it possible for each of the required outputs to select a value for the manipulated variables which will deliver the required output?

Brockett and Mesarovic[6] also define the concepts of Functional Controllability, Asymptotic Controllability, and Pointwise Controllability.

For our purposes we will require only a slight variant of Pointwise Controllability. For Chapters 3 and 4 we require the following Definition D-1.

The system modeled by the state equation $\bar{y} = A\ \bar{m} + \bar{\epsilon}$ is "pointwise controllable" $\longleftrightarrow$ for every point $\bar{y}_0 - \bar{\epsilon}_0$ in the output space $\underline{Y}$ it is possible to select at least one point $\bar{m}_0$ in the space M of manipulated variables such that $A\ \bar{m}_0 = \bar{y}_0 - \bar{\epsilon}_0$.

Several things become apparent. First, if the number of components of $\bar{m}$ is less than that of $\bar{y} - \bar{\epsilon}$ the system cannot be pointwise controllable. If the number of components $\bar{m}$ is greater

than or equal to the number of components of $\bar{y} - \bar{\epsilon}$, we must examine the matrix A. If A is square and non-singular then, of course, the system is pointwise controllable.

What occurs geometrically if the system is not pointwise controllable? For the answer to this question consider the system equation

$$
\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}^T m \tag{D.1}
$$

where in this instance $A = \begin{bmatrix} a_1 \cdots a_n \end{bmatrix}$ and m here is a scalar quantity. We can regard (D.1) as expressing the $y_i - \epsilon_i$ parametrically in terms of the variable m. Hence (D.1) constitutes the equation of a straight line in the output space $\underline{Y}$. Whereas if the system were pointwise controllable we could reach any point of $\underline{Y}$ with some $\bar{m}$ here we can reach only those points on the straight line (D.1) by manipulating m.

For Chapters 5 and 6 we will employ the following concept.

Definition D-2. We will define the system modeled by the equations $\bar{y}_{t+1} = A \bar{y}_t + \bar{m}_{t+1} + \bar{\epsilon}_{t+1} \qquad \bar{y}_o = \bar{c} \qquad t = 0, 1, \ldots$ to be "functionally controllable" $\longleftrightarrow$ every possible sequence of values definable in the output space $\underline{Y}$ can be realized by the selection of at least one sequence in the space of manipulated variables $\underline{M}$.